U.S. Army Research Institute for the Behavioral and Social Sciences

Research Report 1942

The Retention of Digital Skills Following Distributed and Traditional Training

Gregory A. Goodwin & Jennifer S. Tucker
U.S. Army Research Institute

Richard L. Wampler
Northrop Grumman Corporation

Amanda N. Gesselman
Columbus State University
Consortium Research Fellows Program

Vanessa Johnson
Auburn University
Consortium Research Fellows Program

April 2011

U.S. Army Research Institute for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

Barbara A. Black

BARBARA A. BLACK, Ph.D.
Research Program Manager
Training and Leader Development Division

Michelle Sams

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract for the Department of the Army

Northrop Grumman Corp.

Technical review by

Brian T. Crabb, U.S. Army Research Institute
John Lipinski, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Research Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) April 2011	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) July 2007 – Aug 2009
4. TITLE AND SUBTITLE The Retention of Digital Skills Following Distributed and Traditional Training		5a. CONTRACT OR GRANT NUMBER W74V8H-04-D-0045, DO#0019 5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Gregory A. Goodwin, Jennifer S. Tucker (U.S. Army Research Institute), Richard L. Wampler (Northrop Grumman Corporation), Amanda N. Gesselman (Columbus State University), Vanessa Johnson (Auburn University).		5c. PROJECT NUMBER A790 5d. TASK NUMBER 326 5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Institute for the Behavioral and Social Sciences ARI-Ft Benning Research Unit P. O. Box 52086 Fort Benning, GA 31995-2086 Northrop Grumman Corporation 3565 Macon Road Columbus, GA 31907		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences ATTN: DAPE-ARI-IJ 2511 Jefferson Davis Highway Arlington, VA 22202-3926		10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Research Report 1942

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Contracting Officer's Representative and Subject Matter POC: Gregory A. Goodwin

14. ABSTRACT *(Maximum 200 words)*:
As digital systems proliferate in the Army, there is a need for commanders to understand how to best maintain these critical warfighter skills. In the present report, skill retention for the Force XXI Battle Command Brigade and Below (FBCB2) digital system was investigated following traditional face-to-face training and distributed learning (dL) training. This second type of training was examined because little is known about training digital skills using this method. Operator skills were measured immediately and eight weeks following the training. There were no differences in overall performance between the dL and traditional students at baseline suggesting this system can be trained effectively in a dL environment. Both groups also showed similar rates of forgetting after the eight-week retention interval. Overall performance declined significantly from 71% of steps correct at baseline to 62% correct eight weeks later. Characteristics of the digital system and of the participants contributed to forgetting. These findings indicate that FBCB2, and presumably other digital systems can be effectively trained in a dL environment. Additionally system and individual characteristics that contribute to skill decay were identified and can be used to both improve system training and system design.

15. SUBJECT TERMS
FBCB2, Force XXI Battle Command Brigade and Below, command and control, digital skills, skill retention, ABCS, digital training, distributed learning, blended learning

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	67	Ellen Kinzer Technical Publication Specialist 703-545-2445

Research Report 1942

The Retention of Digital Skills Following Distributed and Traditional Training

Gregory A. Goodwin & Jennifer Tucker
U.S. Army Research Institute

Richard L. Wampler
Northrop Grumman Corporation

Amanda N. Gesselman
Columbus State University
Consortium Research Fellows Program

Vanessa Johnson
Auburn University
Consortium Research Fellows Program

ARI – Fort Benning Research Unit
Scott E. Graham, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

April 2011

Army Project Number 622785A790	**Personnel, Performance and Training Technology**

Approved for public release; distribution is unlimited.

ACKNOWLEDGEMENT

This research would not have been possible without the substantial effort of many “behind the scenes” people. The expertise provided by the FBCB2 trainers at Forts Hood and Benning was crucial in identifying the items to be tested and then verifying that the detailed steps of each procedure were properly captured in our test. Their insightful comments and precious time were of immeasurable value. Of special note are Alfredo Lezcano and Scott Hardcastle who once again generously provided critical subject matter expertise.

The authors also are indebted to LTC Thomas Cole, Program Officer, Battle Command Training Center (BCTC), Camp Dodge, Iowa. LTC Cole generously granted access to classes and provided insightful feedback on the research findings. Special thanks also go to Frederick Anderson and Adam Mittower, FBCB2 instructors at the BCTC, Camp Dodge, IA. These instructors generously allowed us to collect data from students during their classes. We also thank Jeffrey Holle who helped coordinate the data collections and Eric Soults who provided feedback on the research findings.

The authors also are grateful for the many Active duty and National Guard Soldiers who among their many duties, graciously gave their time and expertise to this project.

At various times throughout the project many Northrop Grumman personnel made significant contributions to assist with administering tests and arranging for Soldiers to return for retesting. Those who deserve special mention are Chris Strauss at Fort Hood, Greg Burnett at Fort Riley, and Michael Dlubac at Fort Benning. Acknowledgement is especially due to Nancy Riffe and Diadra Swinson of Northrop Grumman who compiled numerous FBCB2 system screen shots, developed detailed graphics and storyboards for each of the items to be tested, and then verified test accuracy. Another important contributor was Sharon Meyers of the Army Research Institute’s Occupational Analysis Office who programmed the storyboards into the automated data collection instrument and patiently made revisions until the test was ready. Without the contributions of all of these people, this project would not have been possible, and to all of them we offer our sincere appreciation.

THE RETENTION OF DIGITAL SKILLS FOLLOWING DISTRIBUTED AND TRADITIONAL TRAINING

EXECUTIVE SUMMARY

Research Requirement:

The proliferation of digital command and control systems on the modern battlefield places a growing training requirement on Soldiers to acquire and maintain these skills. This research continues a program investigating the retention of digital skills. Prior research in this program has shown slight, though significant forgetting of Force XXI Battle Command Brigade and Below (FBCB2) skills over an 8 week retention interval. However, that research was done with officers following familiarization training. In the current effort, we wanted to examine recall following the Army's standard FBCB2 operator course (a 40h course typically attended by junior enlisted Soldiers) to better understand skill retention in the typical user of this system.

Two modes of training were also investigated: traditional face-to-face training and distributed learning (dL) training. With the increased demand to develop and maintain highly proficient system operators, some Army units have utilized dL technologies to train digital skills. Because dL instructional environments have unique training challenges and little is known about the effectiveness of training digital skills using dL, there is a critical need to know whether Soldier performance following dL instruction differs from traditional (face-to-face) classroom instruction.

Procedure:

To measure skill retention, a test of FBCB2 skills was developed that could be administered on a computer either locally or over the internet. Participants were presented with screenshots of the FBCB2 system and were asked to answer multiple choice questions regarding actions they would take on the system (e.g., click a button, choose an option, type information into a field). Questions pertained to procedures or tasks that could be performed on FBCB2. In some cases, Soldiers only had to indicate the first step they would take starting at the operations screen and in other cases they had to complete a series of steps.

Participants from the face-to-face training were Active duty Soldiers (n=80) and from the dL training were National Guard Soldiers (n=136). Of those who took the baseline test, 32 (24%) from the dL group and 31 (39%) from the traditional group returned to take the retention test.

Findings:

Overall, participants from both training environments performed similarly on the baseline test suggesting that dL training is as effective as traditional face-to-face training for FBCB2 operator training. When looking at retention of skills, participants performed 71% of the steps

correctly at baseline. This declined significantly to 62% eight weeks later. Forgetting occurred at the same rate regardless of the initial mode of training.

Item characteristics that contributed to forgetting were investigated by identifying common characteristics of well-recalled questions and of poorly-recalled questions. Well-recalled questions were logically linked to the overall question, were verification steps such as “are you sure” or “ok”, or were item completion steps such as “close” or “apply.” Vague or misleading cues in the system contributed to poor recall of some steps.

Utilization and Dissemination of Findings:

This research should be of use to several audiences. First, leaders should know that FBCB2 skills are quickly forgotten. By eight weeks following training, Soldiers had shown significant levels of forgetting. However, this does not mean that they will need refresher training every eight weeks unless leaders want Soldiers to maintain high levels of readiness at all times. It does indicate that refresher training should be scheduled only a few weeks prior to when Soldiers will need to demonstrate their proficiency. Leaders should also be aware that Soldiers with more general computer experience will retain FBCB2 skill better than those without that experience.

Second, Army leadership should know that FBCB2 skills can be effectively trained in a dL environment. It is important to note that the dL environment used by the National Guard Soldiers was well equipped. It allowed instructors to view all of the students’ actions on their individual systems, allowed the students to see the instructor’s system, and allowed students and instructors to see each other and interact verbally. The important point here is that these findings should not be over-generalized to indicate that any dL training is as good as any face-to-face training.

Finally, training developers and instructors in digital classrooms should know which procedures are poorly recalled so that training and memory aids can be developed to improve retention of those areas. Memory aids should be developed to overcome system cues that are especially vague or confusing.

These findings have been briefed to course developers at the Battle Command Training Center at Camp Dodge, IA and was presented at the 2010 International Interservice Training and Simulation Conference.

RETENTION OF DIGITAL SKILLS FOLLOWING DISTRIBUTED AND TRADITIONAL TRAINING

CONTENTS

Page

Page

LIST OF TABLES

Page

LIST OF FIGURES

The Retention of Digital Skills Following Distributed and Traditional Training

To enhance force capabilities, the U.S. Army has been fielding Army Battle Command System (ABCS) equipment since the mid-1990s. This family of digital command, control, and communication systems is spreading advanced network-based capabilities throughout the force and, increasingly, to lower echelons. The most widely used ABCS system is the Force XXI Battle Command Brigade and Below (FBCB2). This system is found in most vehicles and provides navigation, situation awareness, and messaging capabilities. The remaining ABCS systems make up a suite of battle staff tools for accomplishing command and control functions in tactical operations centers. The battle staff systems include the Maneuver Control System (MCS), All Source Analysis System (ASAS), Advanced Field Artillery Tactical Data System (AFATDS), Command Post of the Future (CPOF), and Battle Command Sustainment Support System (BCS3). Table 1 lists the primary ABCS components for the brigade echelon and below.

Table 1
Principle Components of the Army Battle Command System (Brigade and Below)

Component	Role
Force XXI Battle Command Brigade and Below (FBCB2)	Maneuver forces tool for command, control, communication, and navigation; feeds common operational picture
Maneuver Control System (MCS)	Principal staff tool for planning and controlling maneuver operations; primary source of friendly picture
All Source Analysis System (ASAS)	Principal staff tool for planning and controlling intelligence activities; primary source of enemy picture
Advanced Field Artillery Tactical Data System (AFATDS)	Integrated tool for planning and controlling indirect fires; primary source of fire support picture
Command Post of the Future (CPOF)	Dynamic visualization tool that supports collaborative decision-making and planning
Battle Command Sustainment Support System (BCS3)	Principal staff tool for planning and controlling sustainment operations; integration engine for logistics

Digital systems on a wireless battlefield network allow commanders and leaders to rapidly develop a common view of the battlefield so they can make decisions faster and disseminate messages, orders, and overlays to their subordinates. Although every Army division employs ABCS systems, many units struggle to leverage the full potential of their networked capabilities (Clark, 2005). Warfighting potential is lost because of factors such as recurring hardware and software upgrades, personnel turbulence, and decay of digital skills over time. While some of these factors are beyond the control of unit commanders, providing digital training programs to sustain operator proficiency is well within a commander's purview. Knowledge of how to optimally train and sustain digital proficiency is essential to develop highly proficient ABCS operators and maintain their proficiency.

A number of mostly anecdotal reports suggest that digital skills are quickly forgotten (see review in Goodwin, 2006). If this is true, then digital systems come with a heavy training requirement. Commanders and unit trainers must make time and other resources available for Soldiers to maintain these skills. To develop cost-effective training programs it is necessary to gain a better understanding of which digital skills Soldiers forget the most and the factors that contribute to these performance decrements. Thus, one purpose of the present research is to examine the percentage of Soldiers who remembered specific tasks eight weeks following an intensive FBCB2 operator class so that leaders and trainers can better plan and design training for this system.

Retention of Digital Skills

Digital skills are discrete, multi-step procedures (e.g., navigation through a series of menus and submenus to set parameters and execute commands). They are predominantly cognitive tasks, though they require some motor input through a touch screen, keyboard, or pointing device, and they range from the simple to the complex. Research on skill retention shows that discrete procedural tasks, like digital skills, are more perishable than continuous procedures (e.g., riding a bicycle) or declarative knowledge (Adams, 1987). From this perspective, the anecdotal evidence of the fast rate of decay of digitals skills is not surprising.

Although there is anecdotal evidence that digital skills are easily forgotten, there is little empirical evidence documenting the extent of the decay and the specific tasks that are most likely to be forgotten (Goodwin, 2006). An early effort to measure digital skill retention in an Army system was an experiment by Sanders (1999) in which 28 Soldiers were trained to perform tasks using the inter-vehicular information system (IVIS). This system was a forerunner to FBCB2 that provided messaging and navigational capabilities to the M1A1 Abrams Tank crew. In that experiment, participants received training and then performed a series of eight overlay tasks and six report tasks. Participants had to reach a criterion of correctly performing three tasks of each type. At the conclusion of baseline training, 22 participants had reached the criterion on the overlay tasks and 23 had reached the criterion on the report tasks. Following a 30-day retention interval, only 48% of the overlay-skilled group and only 77% of the report-skilled group were able to re-achieve the criterion. Both of these were statistically significant declines.

In a more recent experiment, the retention of FBCB2 skills was measured in 54 officers following a two-day familiarization course (Goodwin, Leibrecht, Wampler, Livingston, & Dyer, 2007). All 54 participants were students attending the Infantry Captains' Career Course (ICCC) at Fort Benning, GA in 2006. Participants were given a hands-on test immediately following training and again 28 days later. Retention was measured differently in this experiment than in the Sanders (1999) report. Participants had only one chance to perform 13 different hands on tasks. At baseline, participants performed an average of 72% of the tasks correctly and 28 days later they performed an average of 62% correctly. This was a significant yet more modest decline than was observed by Sanders (1999). Significant declines in performance occurred for only 3 of the 13 tasks and declines in the percent of the sample able to perform those three tasks ranged from 19% to 24%.

Given that the IVIS system was a forerunner of the FBCB2 system, it might be tempting to conclude that the more recent system was better designed to make these skills more resistant to decay. However, such a conclusion would be difficult to support given the number of methodological differences between the Sanders (1999) effort and the Goodwin, Leibrecht, Wampler, Livingston, & Dyer(2007) experiment. Thus, one purpose of the present research is to confirm and further extend our understanding of FBCB2 skill retention by using a more sensitive measure of skill decay. Additionally, in the present experiment, we sought to examine the performance of Soldiers who typically operate this system in theater, namely enlisted Soldiers, rather than officers as was done in the prior skill retention effort. Finally, in the present experiment, skill retention was examined after the standard 40-hour operator course rather than the two-day familiarization course of the prior effort.

Yet another factor examined in the present research was the mode of training. The Army has increasingly been using the Internet to deliver training to students at remote locations from the instructor. This mode of training, called distributed learning (dL), helps to reduce costs associated with transporting students to a common classroom and reduces Soldiers' time away from their units and families. It is likely that the Army's use of this mode of instruction will continue to expand so it is important to understand skill retention in a dL environment.

Training Digital Skills in Distributed Classroom Environments

In a traditional digital training classroom, students are seated at a computer running the tactical software. Students view the instructor's system on large display screens at the front of the classroom. Instructors use the large display to demonstrate actions on the system while the students mimic those same actions on their own systems. To help individual students that have difficulty, the instructor or assistant instructor can look over the shoulder of the student and provide individual feedback. Until somewhat recently, these types of interactions would have only been possible in a face-to-face classroom.

The widespread adoption and availability of digital learning technologies have made synchronous, or high-fidelity, interactions in a distributed environment possible (Bonk & Graham, 2006). However, few examples of this type of instruction can be found in military settings. Thus, there is a need to determine effective methods for using dL techniques to train different echelons and skills.

One notable exception, and of importance to the present research, is an Army National Guard course that has blended instructional approaches to train FBCB2 operator skills. The course is designed for distributed instruction such that the Soldiers are in computerized classrooms in their home states, remote from the instructor, who is located at the Battle Command Training Center (BCTC), Camp Dodge, IA. The instructors use video-teletraining (VTT) and sophisticated computer software that emulates the FBCB2 system to conduct the digital skills training. As such, the instructors employ a variety of techniques to cover the content, maintain the students' motivation, and sufficiently address students' problems and concerns.

Prior research investigating the different types of training techniques used by the National Guard dL instructors as well as students' reactions to this type of instruction found that the dL instructors were able to use the technology to teach FBCB2 operations as effectively as the traditional courses (Tucker, McGilvray, Leibrecht, Strauss, Perrault, & Gesselman, 2009). The prior investigation also recommended that instructors leverage student strengths, emphasize problem-centered instruction, and leverage training aids (Tucker et al., 2009).

It is interesting to note that the assumption that traditional face-to-face (F2F) training always leads to better student outcomes has *not* been supported by recent meta-analytic findings. That is, meta-analysis with rigorous standards for study inclusion (i.e., only those studies who met certain methodological criteria were included in the analyses) found that students who took all or part of their instruction online performed better than those in F2F instructional environments (Means, Toyama, Murphy, Bakia, & Jones, 2009). The results, showed that the higher performance levels attained by the online students was probably due to their instructors promoting more time on task throughout the courses. Thus it was a property of the instruction and not the delivery medium that improved performance.

The meta-analyses cited above focused primarily on academic courses which may not generalize to a software operator course like the FBCB2 course. Therefore, another goal of this effort was to evaluate the effectiveness of training this system in a dL environment. Student performance across the dL and traditional/F2F instructional environments was compared by examining the retention of FBCB2 skills immediately following the course and again eight weeks after training. The BCTC at Camp Dodge, IA provides a unique opportunity to address this goal. Thus, we selected the National Guard Soldiers who received FBCB2 training as the dL sample for the present research. It is important to note that both the dL and F2F instructors followed the FBCB2 Program Manager's Brigade Battle Command Program of Instruction for the classes included in this research. The training materials were therefore presented similarly for both of the training environments. Although direct comparisons of the instructors who taught the classes for the present research cannot be made, past observations of both instructional environments indicated that instructors use similar training techniques across these classes (see results presented in Leibrecht, Goodwin, Wampler, & Dyer, 2007 and Tucker, McGilvray, Leibrecht, Strauss, Perrault, & Gesselman, 2009).

Factors Contributing to Skill Decay

Research has shown that there are three general categories of variables that impact skill retention: procedural, individual, and task variables (Hagman & Rose, 1983). Procedural variables reflect the properties of the training and testing. Examples of these variables include massed vs. distributed training, training to proficiency or to mastery, and duration of the retention interval. Examples of individual variables include the aptitude of the individual and whether or not the individual has certain background knowledge or expertise. Examples of task variables include the number of steps involved in the task, the complexity of the steps, and whether the task is continuous or discreet.

In the present effort, we examined task variables that might impact retention by having non-FBCB2 users (i.e., individuals with no FBCB2 training or experience) perform the FBCB2 tasks. Our reasoning was that without training on the system, individuals would have to rely on the system's cues to complete the tasks. It was expected that tasks that were easily completed by non-FBCB2 users would show higher rates of recall by the Soldier sample than tasks that were not easily completed by non-FBCB2 users. In addition, we did a *post hoc* characterization of individual questions that were recalled at high and low rates. Examining the way system cues contribute to recall and skill decay provides trainers and training developers with a better understanding of where to focus their efforts and provides system developers with guidance on how to best improve the user interface.

Present Research

In summary, the present research had three primary goals. The first goal was to examine FBCB2 skill retention following the standard 40-hour operator course and to do so using a more sensitive measure of skill decay than was used in prior skill retention research. The second goal was to compare student outcomes from two different classroom environments – traditional and dL – to offer Army leaders and training developers insight into the effects of different instructional approaches on student outcomes. The third goal was to examine some of the task variables that might contribute to skill decay on the FBCB2 system.

Method

Participants

A total of 216 participants took the baseline test. Of those, 136 (63%) completed dL training and 80 (37%) completed traditional training. Of those who took the baseline test, 32 (24%) from the dL group and 31 (39%) from the traditional group took the retention test. The different return rates among active duty (traditional/F2F) and National Guard (dL) participants were most likely due to Active Duty Soldiers being given time to return to the test site during the duty day (unless they had other priorities) to complete the retention measure. The National Guard Soldiers were asked to voluntarily complete the retention measure on their own time. Active Duty Soldiers were drawn from six traditional classes at Fort Hood and Fort Riley between August 2008 and April 2009. National Guard Soldiers were drawn from 12 different dL classes from March to October of 2009.

To determine that the sample who took the retention measure did not differ substantively from the larger, baseline sample, z-test comparisons were made between participants who did and did not complete the retention measure. The only differences between the baseline only and retention samples were for the dL sample. When rating the duty position in which they would be most likely to use FBCB2, participants in the dL retention sample were more likely to say they would do so as a vehicle commander than the dL baseline only sample (19% vs. 8%). The dL retention sample also had significantly higher self-ratings of computer proficiency (3.7 vs. 2.8) on an 8-point scale than the dL baseline only sample. Finally, the dL retention sample was more likely to have used MCS while deployed than the dL baseline only sample (56% vs. 10%). , There were no differences between the baseline only and the retention samples on rank, likelihood to use FBCB2 when deployed, duty position in which FBCB2 would be used if deployed, self-ratings of proficiency on any ABCS system, or prior combat experience using any ABCS system. With the possible exception of general computer experience, the differences observed would not therefore be expected to impact performance on the skill retention measure. The demographic data in Tables 2-4 summarize responses from all participants. Additional demographic summaries are presented for the baseline and retention samples in Appendix A.

In addition to comparing the baseline only and retention samples, it was necessary to compare the samples from the two training environments. As can be seen in Table 2, the traditional sample was more junior than the dL sample. Sixty percent of the traditional sample was made up of privates and specialists compared to only 33% of the dL sample. On the other hand, 53% of the dL sample was made up of sergeants and staff sergeants compared to 25% of the traditional sample. The greatest percentage of the traditional sample was in the combat arms branch whereas the greatest percentage of the dL sample was in the combat support branch. Expected primary duty positions reflected differences in the ranks of the two groups, with the dL sample being more likely to expect to serve in company and above echelons and the traditional sample being more likely to serve at the platoon level.

Table 2
Demographic Data for Participants

Question	% of Sample Traditional	dL
Current Grade/Rank:		
E3 (Private)	30	5*
E4 (Specialist/Corporal)	30	28
E5 (Sergeant)	15	29*
E6 (Staff Sergeant)	10	24*
E7 (Sergeant First Class)	5	5
E8 (First Sergeant/Master Sergeant)	0	5
O1/O2 (Lieutenant)	9	2*
O3 (Captain)	0	2
Branch:		
Maneuver/Combat Arms	44	37
Maneuver Support/Combat Support	24	46*
Sustainment Support/Combat Service Support	32	17*
Likelihood of using FBCB2 if deploying immediately:		
Very unlikely	17	10
Somewhat unlikely	9	16
Somewhat likely	37	46
Very likely	38	28
Role/position in which FBCB2 would be used if deploying immediately:		
Unknown	35	28
Primary Operator for a leader	15	8
Section Leader/Squad Leader	22	24
Vehicle Commander (other than Leader/Commander)	8	10
Platoon Leader/Platoon Sergeant	14	4*
Company/Headquarters Support Element	4	12*
Company/Troop Commander	0	1
Staff Officer/Noncommissioned Officer Battalion or Brigade Tactical Operations Center	3	15*

Note: * $p < .05$. Comparisons based on z-test for equal proportions in independent samples.

Participants were also asked to rate their proficiency on five digital systems (Table 3). The only system on which differences were seen was FBCB2, with the traditional students rating themselves at higher levels of proficiency than the dL students. Additionally, a small but significantly larger percentage of dL students had never used FBCB2 as compared to the traditional sample.

Table 3
Self-Ratings of Proficiency on Five Digital Systems

	Percent of group							
	Never Used		Basic		Medium		High	
System	Traditional	dL	Traditional	dL	Traditional	dL	Traditional	dL
FBCB2	1	12*	13	45*	72	38*	14	5*
ASAS	79	73	10	19	11	8	0	1
AFATDS	85	85	9	11	6	2	0	3
MCS	76	72	19	19	5	7	0	2
CPOF	89	87	4	9	8	3	0	2

Notes. Comparisons based on z-tests for equal proportions in independent samples. Self-rating categories are described in the text.
*p < .05.

The higher self ratings of proficiency on FBCB2 by the traditional participants are not too surprising in light of self-reports of training and experience (Table 4). A significantly higher percentage of participants in the traditional sample had FBCB2 training (in addition to the training received just prior to the data collection). Additionally, a higher percentage of traditional participants had used FBCB2 on deployment, though the difference was not significant. Training and experience on the other digital systems was comparable across the two groups. Participants also indicated the amount of training received during the eight-week retention interval for the present research. Of the 32 dL participants who took the retention measure, 31 responded to this question. Only five Soldiers in the dL classes received any training (M = 1.8 hours). All 31 of the traditional participants responded to this question. Only six received any training (M = 2.1 hours). This difference was not statistically significant

Table 4
Training on Five Digital Systems by Users of Each System

	Percent of System Users			
	Training		Use While Deployed	
System	Trad.	dL	Trad.	dL
FBCB2	37*	18	27	16
ASAS	9	8	0	0
AFATDS	54	23	8	0
MCS	60	54	5	21
CPOF	56	61	13	11

Note. Comparisons based on z-test for equal proportions in independent samples.
*p < .05.

Despite the fact that the traditional sample had more training and experience, these two samples also had much in common. In both samples, enlisted ranks made up at least 90% of the sample. In addition, about three quarters of each sample expected to use FBCB2 when deployed saying they were either somewhat likely or very likely to use it. Everyone in both groups had used a computer before. Finally, self ratings of proficiency, prior training, and experience indicated that of all ABCS systems, both samples were, by far, most familiar with FBCB2.

Although it might have been preferable from an experimental standpoint to have matched all background variables from both dL and traditional samples, it was not feasible to be so selective in our sampling. For the traditional sample, it was extremely challenging to have Soldiers come back to be retested given unit deployment schedules. Thus, the approach taken in the present research has been to interpret differences between these two training samples in light of the background and training differences observed.

Measures

Development of test items. In developing the test, a goal was to include a set of procedures that varied in terms of frequency of use and complexity. Data on FBCB2 frequency of use and criticality to mission success were available from a recent research project with 636 combat veterans with Operation Iraqi Freedom (OIF) experience (Bink, Wampler, Goodwin & Dyer, 2009). The procedures examined in that project were included in a list of 53 procedures that were considered for testing in the current skill retention effort. From this list, procedures were chosen that were frequently used (e.g., creating and sending a free text message) and infrequently used (e.g., attach an overlay to an operations order). Difficulty was reflected by the number of steps to complete the procedure, and procedures were chosen that required relatively few steps to complete (e.g., create a message folder – 5 steps) and many steps (e.g., clear logs and queues – 11 steps). Moreover, it was critical that the measure include only procedures that were taught in the 40-hour operator courses. The procedure list was reviewed by both dL and traditional course instructors to ensure that the procedures were taught in the course.

To keep the test to a reasonable length (approximately 1-hour to complete) while still testing a variety of different procedures, it was not possible to ask participants to perform all steps of all procedures. Therefore, three types of performance measures were developed: operations screen questions, full procedure items, and partial procedure items (see Table 5). For the operations screen questions, participants were presented with a view of the primary map display (the operations screen) and asked to indicate the first step needed to initiate each of 14 procedures. For the full procedure items, participants had to perform all steps of seven procedures. For the partial procedures, participants performed only a subset of the steps for five procedures.

To help distinguish between the different types of performance measures, the term “question” is used in this report to refer to each individual multiple-choice item. The question is the most basic unit of performance measurement. For each question, the participant had a multiple choice question that referenced an FBCB2 screenshot. The answers corresponded to the possible actions that could be taken on that system screen. There were between 100 and 105 individual questions. The range of questions is due to the fact that not all branches of each multi-step item had equal numbers of questions. Thus the number of questions answered by each participant depended on choices made at some branchpoints. The term “procedure” is used to refer to the set of questions that make up a multi-step item. There are two types of procedures: full and partial. Finally, the term “item” is used as a generic term to refer to all three types of performance measures: operations screen questions and full and partial procedures.

In addition to measures of procedural knowledge, there were five questions that examined declarative knowledge of the system. The five knowledge-based questions centered on the participants' knowledge of the meaning of specific information or screen prompts visible on a variety of FBCB2 screens. The analysis and discussion of the knowledge questions are found in Appendix B.

Table 5
Items Included in FBCB2 Assessment

Item Name	# Steps Tested
Operations Screen Items	
Check GPS software to troubleshoot GPS problem	1
Manually place your vehicle icon on the map	1
Enter the MEDEVAC (medical evacuation) call sign and voice net frequency	1
Create a periodic reminder	1
Create an address group	1
Assign message to quick-send button	1
Create and save a position report	1
Activate driver's display for a route	1
Use the circular line of sight tool	1
Display an overlay message	1
Create and send an NBC1 report	1
Create and send a Mayday report	1
Transmit combat platform status/SITREP (situation report)	1
Edit a location folder	1
Whole Procedure Items	
Create a message folder	5
Set default addressing for a SPOT (size activity location time) Report	8
Clear logs and queues	11
Create, save and send a SPOT report	7/4
Create and send a route	11
Set screen to display all enemy units and only current, friendly units	6/7
Show a specified vehicle on the display	6
Partial Procedure Items	
Create and save an overlay object group	12/11
Create and save a Combat Services Support overlay	11
Attach an overlay to an OPORD (operations order)	6
Display a satellite image on SA display	3
Display MGRS (military grid reference system) gridlines on map	4

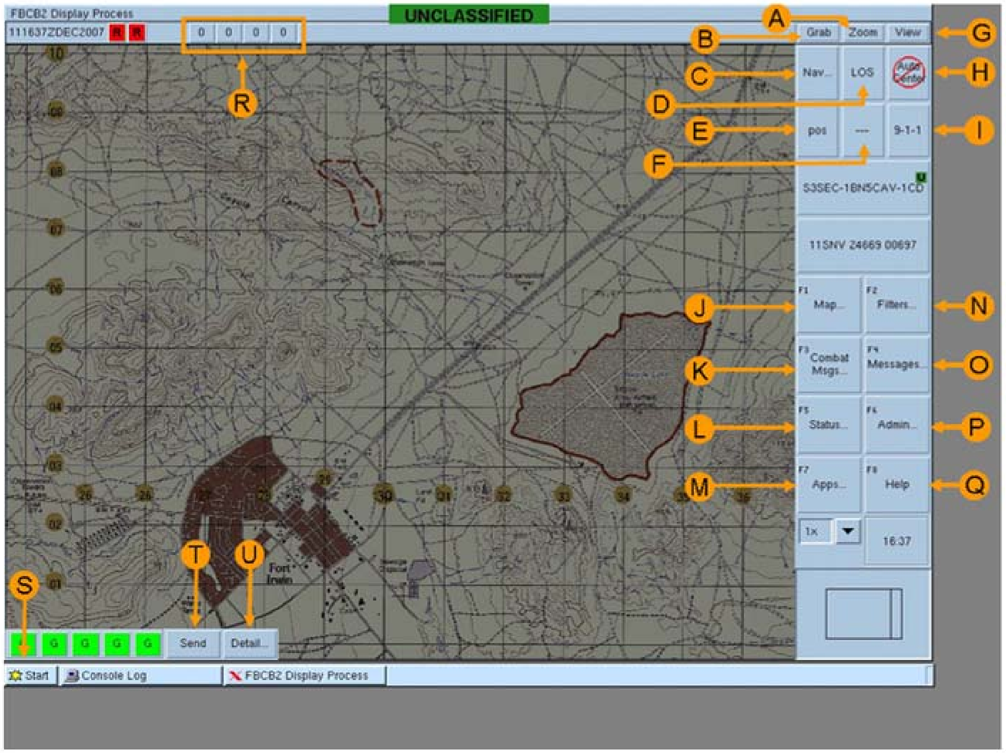
Note: Multiple numbers of steps are indicated where branching occurred and the alternate branches had different numbers of steps.

For all procedural and knowledge questions, participants were presented with a screenshot from FBCB2 with overlaid letters indicating all possible valid responses. A sample of a procedural question (an operations screen question) is provided in Figure 1. Beneath the screenshot, the question and all valid responses were presented. A response was not registered until the participants clicked on the “next question” button. For the operations screen questions, participants had the option to go back and change previous answers by clicking on a “previous question” button. However, in the whole-procedure and partial-procedure sections, this option was removed because the correct response to each question was displayed in the subsequent step. So, for those two sections, all first responses were final. A summary of all the test items including sample screen shots is found in Appendix C.

Development of the data collection instrument. There were 127 individual questions with screenshots in the skill retention instrument; screen images, questions, and the possible answers were assembled into a packet of storyboards. With these storyboard packets as a guide, the measure was developed using InterForm® software (Rao, 2001) that could be administered over the Internet using a standard web browser. Alternatively, for computers without Internet access, the measure was administered with an executable file run from a portable storage device. The measure was identical regardless of which mode of administration was used. Specifically, FBCB2 screenshots depicting the steps of a procedure were presented to the participants. For each step, participants indicated what action needed to be performed (e.g., clicking a button, setting an option, typing required text; Figure 1).

This approach offered several advantages over testing participants on the operational software as was done in a prior skill retention effort (Goodwin et al., 2007). First, the computerized measure made it possible to know the percent of all steps performed correctly, making it a more sensitive measure of recall. When using the operational software in the previous effort, participants could perform procedures using a trial and error approach. Because completion of the procedure was the criterion, it was not possible to distinguish between a participant who knew exactly how to perform the procedure without error and a participant who completed the procedure making many errors along the way. Second, the computerized measure made it possible to identify which steps were most likely to be forgotten, something that was not possible using the operational software. Finally, the measure used in the present effort did not have a help function as found in the operational software. The help function in the operational software may have masked the true skill decay levels in the previous effort.

The chief disadvantage of using the computerized measure was that it was impossible to allow for all possible pathways to complete each procedure. To allow for multiple ways to complete a given procedure, several of the questions had branch points where a participant’s response determined alternate pathways through a procedure. As some procedures had multiple ways to accomplish certain steps, branch points were developed for the most typical ways of performing the steps. Thus, sometimes a choice that would be considered correct was not followed by the screenshot that reflected that choice. When scoring these responses, participants were given credit for a correct response.



2. Your GPS is not operating at the optimum level and you need to troubleshoot the problem. The hardware has been checked and there were no problems. You need to check the software. **What is the first step?**

○ A - Zoom	○ D - LOS	○ G - View	○ J - F1 Map	○ M - F7 Apps...	○ P - F6 Admin...	○ S - Start
○ B - Grab	○ E - pos	○ H - Auto Center	○ K - F3 Combat Msgs...	○ N - F2 Filters...	○ Q - F8 Help	○ T - Send
○ C - Nav...	○ F - ---	○ I - 9-1-1	○ L - F5 Status...	○ O - F4 Messages...	○ R - 0 0 0 0	○ U - Detail...

Previous Question   Next Question

Exit & resume testing later

Figure 1. Example of an operations screen question on the computerized test

Procedure

Traditional training sample. Active-duty Soldiers received training at Forts Hood and Riley. Training occurred in classrooms equipped with personal computers running the operational FBCB2 software. The instructor's system was displayed on a large screen at the front of the classroom. The active-duty participants were trained in classrooms with about 20 – 30 students. Training typically involved demonstration and explanation of procedures with opportunities for the students to practice the steps on their own computers. Instructors and assistant instructors could walk around the classroom to view the students' progress and address questions. These training techniques are described in detail in Leibrecht, Wampler, Goodwin, and Dyer (2007).

Soldiers were administered the FBCB2 measure immediately after completing the final requirement for the 40-hour FBCB2 operator's course and again 8 weeks later. Participants provided contact and chain-of-command information so that they could be contacted prior to the administration of the retention measure. The baseline and retention measures were administered on personal computers. For the retention measure, Active Duty Soldiers returned to the digital system training facility where they received their initial operator training. A username and password, provided by the participants at baseline were used to match each participant's data across the two tests. Contact information was destroyed after the participants completed the retention measure.

dL sample. The National Guard Soldiers received training at digital classrooms in their respective states from instructors located at Camp Dodge, IA. Although they were in different locations, the instructors and students could see one another. The dL instructor had a view of each student's computer screen. Similarly, each student could see the instructor's computer screen. The students' and instructors' computers ran software that emulated FBCB2 exactly. That is, unlike the traditional students who used the actual FBCB2 software, the dL students ran software that looked and acted just like the actual FBCB2 software. The training techniques used by instructors in this dL environment were similar to those used by the instructors in the traditional classes. These techniques are described in detail in Tucker, et al. (2009). Moreover, as indicated above, the instructors employed the same POI as the instructors of the traditional classes.

All of the dL data were collected via the Internet using the same baseline and retention measures as the traditional classes. The researchers were present for data collection during the administration of the baseline measure. The Soldiers were asked to log into the survey website and complete the assessment prior to the end-of-course exercise. For the retention measure, taken 8 weeks later, Soldiers were sent an e-mail with the link to the retention measure and the username and password information they provided at the baseline assessment. As with the traditional sample, usernames and passwords were used to match the baseline and retention data for each participant. All personal information (names, e-mail addresses) were destroyed after the retention data were collected.

Non-FBCB2 users. A group of six raters, who had never received FBCB2 training and who had never used FBCB2 prior to the present research, completed a paper and pencil version of the retention measure. Participants were instructed to read each question, look at the screenshot and all possible responses and choose the response that seemed most logical. As described above, the reason for including these non-FBCB2 users as participants was to help examine the intuitiveness of the system.

Analyses

Different participant samples were examined to address several questions about training and retention. To examine, proficiency levels immediately following training, the entire baseline sample was examined. To report proficiency levels after the 8-week retention interval, only those individuals who took both the baseline and retention measures were examined. To examine skill decay following training to proficiency, only those individuals who performed all of the steps of a procedure correctly at baseline and who completed the retention measure were examined. This latter sample also was used to identify the steps that were most and least likely to be recalled.

Performance was averaged across all questions, across the set of single-step items (operations screen questions), and across the multi-step items (full and partial procedures). Performance on individual items is reported in Appendix D. Performance on the set of operations screen questions was determined by calculating the mean percent correct of the 14 operations screen items. Performance on the multi-step items was determined by first calculating the mean percent of steps correct for each procedure and then averaging across all procedures. Overall performance was the percent correct of all questions answered by each participant.

In the section examining retention in individuals with perfect baseline performance, the goal was to understand retention rates following training to proficiency. That is, given a group of individuals who all answered a question correctly at baseline, how many still answered it correctly eight weeks later? In this set of analyses, each item on the test was examined independently because none of the participants answered all of the questions correctly at baseline.

Results

Baseline Comparisons

Overall hands-on performance (Multi-step & Operations Screen items). As described above, an overall score of all hands-on items was calculated for each Soldier. This measure was analyzed in a two factor, instructional environment by times tested, analysis of variance (ANOVA). There were no significant main effects of either instructional environment (dL vs. traditional) or times tested (baseline only vs. baseline and retention). However, there was a significant interaction of these two factors, $F(1, 210) = 11.5$, $p < .01$. Those in the dL sample who completed the retention measure performed better at baseline than those who didn't return, $F(1,210) = 10.8$, $p < .01$. Those in the traditional sample who returned to take the retention measure tended to do worse than those who didn't, though this was not a significant decline (Table 6).

Table 6
Baseline Results for Overall Hands-On Test Scores

Instructional Environment	Times Tested			
	Baseline Only		Baseline and Retention	
	M	n	M	n
dL	67%	102	76%	32
Traditional	72%	49	66%	31

Multi-step procedures (whole & partial procedures). Comparisons between the traditional and dL classes were made for the full sample of baseline participants on the multi-step (whole and partial) procedures. As described above, the performance measure for the multi-step procedures was the average percentage of steps performed correctly per procedure.

Each participant's mean score of all multi-step procedures (Table 7) was analyzed in a two factor, instructional environment (dL vs. traditional) by times tested (baseline only vs. baseline and retention), analysis of variance (ANOVA). There were no significant main effects of either instructional environment or times tested. However, there was a significant interaction of these two factors, $F(1, 210) = 13.7$, $p < .01$. The Soldiers in the dL classes who completed the retention measure performed better at baseline than those who did not, $F(1,210) = 10.7$, $p < .01$, whereas Soldiers in the traditional classes who returned for the retention test performed worse than those who did not, $F(1,210) = 4.2$, $p < .05$. Thus, overall proficiency levels appear to be comparable in both samples immediately following training. However, those returning to take the retention test were different than those who did not. These findings may be the result of a self-selection bias in the dL sample such that the higher performers at baseline were the ones who completed the second measure. This could have been more likely in the dL sample as they had to take the retention measure on their own time whereas traditional students were sent to take the retest during the duty day.

Table 7
Baseline Results for Overall Multi-Step Items

Instructional Environment	Times Tested			
	Baseline Only		Baseline and Retention	
	M	n	M	n
dL	67%	102	76%*	32
Traditional	72%	49	66%*	31

* $p < .05$ as compared to baseline only group
Note: These means happen to be identical to those in Table 6.

Operations screen questions. To examine overall differences between the dL and traditional classes, the percent of the 14 operations screen questions answered correctly at baseline was computed for each Soldier. A two factor, instructional environment (dL vs. traditional) by times tested (baseline only vs. baseline plus recall), ANOVA revealed only a significant effect of instructional environment, $F(1, 210) = 18.9$, $p < .01$, with Soldiers in the traditional classes performing better (75% correct) than the Soldiers in the dL classes (62% correct – see Table 8).

Table 8
Baseline Results for Overall Operations Screen Items

	Times Tested			
	Baseline Only		Baseline and Retention	
Instructional Environment	M	n	M	n
dL	60%	102	67%	32
Traditional	74%	49	76%	31

Note. There was a main effect of instructional environment with traditional students doing better than dL students.

In summary, the analysis of the baseline results suggest that the dL and traditional samples performed similarly following training. When looking at different types of items, however, the traditional students performed better than the dL students on the operations screen questions. There was some indication that the dL participants returning for the retention test performed better than those who did not return. As there was no way to control individual choices by dL participants to return and take the retention measure, these differences must be considered when interpreting the retention results.

Retention Results

Overall hands-on performance comparison results. Participants were asked to indicate whether they received any training on FBCB2 following administration of the baseline measure. Of the dL participants, 5 received additional training on FBCB2 and of the traditional participants, 6 received additional training. On average, participants in each group received 2 hours of additional training during the 8 week retention interval.

Table 9
Retention Results for Overall Hands-on Performance

Average of All Multi-step & Operations Screen Items on Test	dL *n*=30		Traditional *n*=31		Average % Decrease
	Baseline	Retention	Baseline	Retention	
Mean Score	75%	64%	66%	61%	8%

As with the baseline comparisons, retention was examined for overall performance. Data were analyzed in a mixed design, two factor (time by instructional environment) ANOVA. The results from the ANOVA indicated only a significant main effect of time with a decline from baseline to the 8-week retention test (71% correct vs. 62% correct), Wilks' $\Lambda = .62$, $F(1,61) = 36.9$, $p < .01$. Table 9 shows both the baseline and retention average percentages for the traditional and dL participants for this subsample. Table 9 also shows the average percent decrease in performance for the overall score across both groups.

Multi-step procedures (whole & partial items). As described in the Method section, 30 Soldiers from the dL sample and 31 from the traditional sample completed the retention measure 8 weeks following the initial assessment. Changes in performance across the 8-week retention interval in these two samples were compared to determine the overall change in proficiency.

For the multi-step procedures, the average percent of steps correct for each item was computed for each of the 12 multi-step procedures. This overall measure was analyzed in a two factor, time by instructional environment, ANOVA. This analysis revealed a significant effect of time, Wilks' $\Lambda = .80$, $F(1, 59) = 14.4$, $p < .01$ and instructional environment, $F(1, 59) = 5.8$, $p < .05$. Overall performance was higher at baseline than at retention (71% correct vs. 65% correct) and dL participants performed better across both times than traditional participants (72% correct for dL vs. 64% correct for traditional). The better performance by the dL participants in this subsample was not too surprising given the finding, reported above, that dL participants who returned for the retention measure, performed better than those who did not return. There was no significant interaction between time and type of training indicating the dL and traditional participants forgot at the same rate overall (Table 10).

Table 10
Retention Results for Multi-Step Procedures

Multi-Step Procedures	dL *n*=30		Traditional *n*=31		Average % Decrease
	Baseline	Retention	Baseline	Traditional	
Mean Score	76%	68%	66%	62%	6%

Operations screen questions. Comparisons between the traditional and dL classes also were made for the operations screen questions. As before, only those participants who responded to both the baseline and the retention measures were included in the analyses. An average percent correct operations screen score was analyzed using a 2 (instructional environment) x 2 (time) repeated measures ANOVA. The results indicated a similar pattern of skill decay for both instructional environments. There was only a main effect of time, Wilks' $\Lambda = .44$, $F(1, 61) = 78.82$, $p < .01$; the main effect for instructional environment and the instructional environment by time interaction were not significant. Table 11 shows both the baseline and retention average percentages for the traditional and dL participants. Table 11 also shows the overall average percent decrease in performance for these items across both groups.

Table 11
Retention Results for Operations Screen Items

Operations Screen Questions	dL *n*=30		Traditional *n*=31		Average % Decrease
	Baseline	Retention	Baseline	Traditional	
Mean Score	67%	51%	76%	53%	20%

Retention Results for Soldiers with Perfect Baseline Performance

In the prior section, Soldier performance was examined for all Soldiers completing the baseline and retention tests regardless of whether they performed procedures correctly at baseline. Thus, the previous section of this report demonstrates the "real-world" expected proficiency of Soldiers immediately and eight weeks following FBCB2 operator training for the items we examined.

To examine retention following training to proficiency, the subsample of Soldiers who performed each procedure perfectly at baseline was examined. In this set of analyses, the sample sizes vary depending on the number of Soldiers who were able to perform any given multi-step procedure/operations screen question correctly at baseline. To maximize the size of these subsamples, data from the two instructional environments were combined (maximum of 61 Soldiers).

Table 12 presents the percentage of Soldiers who correctly answered each multi-step procedure at eight weeks after having performed all the steps correctly at baseline. Wilcoxon tests were used to determine whether declines were statistically significant. In addition, Table 12 lists the average percent of steps completed at eight weeks by those who performed each procedure perfectly at baseline. One sample t-tests were used to determine statistically significant declines in this measure compared to an expectation of perfect performance.

The statistical analyses in Table 12 should be interpreted cautiously as the sample sizes for some of the procedures were quite small. In two cases, *Save an overlay object group with a hostile light infantry unit* and *Attach overlay to OPORD*, the number of participants was too small to perform statistical analyses. In other cases, the small sample sizes made it almost impossible to detect significant changes.

Taken as a whole, it can be seen that there is substantial forgetting over the 8 week retention interval. Fewer than one third of Soldiers with perfect baseline performance were still able to perform 10 of the 12 multi-step procedures at retention. Fewer than half of these participants could perform 7 of the 14 operations screen questions at retention.

It is worth noting that the percent of steps correct for each procedure are substantially higher than the percent of Soldiers able to perform the procedure correctly at recall. For example, for *Set default addressing for a SPOT report*, only 8% of participants who performed correctly at baseline still remembered all the steps 8 weeks later, yet on average participants recalled 80% of the steps.

Table 12
Percent of Perfect Baseline Performers Still Able to Perform Multi-Step Procedures at 8 Weeks

Procedure	n	% Soldiers Correct at recall	*Z*	% Steps Correct at Recall	*t*
Create, save, and send a SPOT Report	39	64*	-3.7	88*	*t*(38)= -3.9
Display MGRS gridlines on your map	20	50*	-3.2	83*	*t*(19)= -3.6
Find a Unit / Platform on the map	6	33	-2.0	72*	*t*(5)= -2.5
Create and send a route	7	29	-2.2	82*	*t*(6)= -2.9
Create and save a Combat Services Support overlay	7	29	-2.2	76*	*t*(6)= -3.0
Set filters to display all enemy units and only current, friendly units	17	24*	-3.6	68*	*t*(16)= -4.2
Clear logs and cues	22	23*	-4.1	81*	*t*(21)= -6.4
Display a satellite image on your SA display	22	22*	-3.6	74*	*t*(21)= -4.5
Create a message folder	25	20*	-4.5	74*	*t*(24)= -6.5
Set default addressing for a SPOT Report	23	9*	-4.6	80*	*t*(22)= -5.9
Attach an overlay to an OPORD	2	0	-1.4	83	--
Create and save an overlay object group	1	0	--	36	--

Notes: Baseline performance was 100% for all participants. Z is the Wilcoxon test statistic used to analyze the percent of Soldiers able to perform the procedure at both timepoints. The *t* statistic is the one sample *t* test comparing the percent steps correct at recall to perfect performance. SPOT- includes size, activity, location, and terrain; MGRS – military grid reference system; OPORD – operations order
*$p < .05$ significant decline from baseline

Table 13 presents the percentage of Soldiers who correctly answered the operations screen questions at eight weeks after having performed them correctly at baseline. Because the baseline performance of this subsample is always 100%, only the retention percent is included in Table 13. Wilcoxon tests were used to determine statistical significance of changes. As can be seen, significant skill decay was observed for all but two of the operations screen questions.

Table 13
Percent of Perfect Baseline Performers Still Able to Perform Operations Screen Questions at 8 Weeks

Question	*n*	% Soldiers Correct at recall	*Z*
Create and send NBC1 report	60	95	-1.7
Create and send Mayday report	59	93	-2.0
Assign a message to a quick send button	52	89*	-2.5
Create and save a new position report	53	83*	-3.0
Display an overlay message that was sent to you	55	69*	-4.1
Transmit your combat platform status / SITREP	35	69*	-3.3
Create an address group	55	64*	-4.5
Manually place your icon / platform on the map	48	46*	-5.1
Use the circular line of sight tool	46	46*	-5.0
Activate the driver's display for a route	38	42*	-4.7
Create a periodic reminder	53	40*	-5.7
Check software to troubleshoot a GPS problem	18	39*	-3.3
Edit a location folder	16	31*	-3.3
Enter the MEDEVAC call sign and voice net frequency	30	23*	-4.8

Notes: Baseline performance was 100% for all participants. Z is the Wilcoxon test statistic. NBC- nuclear, biological, chemical; SITREP – situation report; MEDEVAC – medical evacuation; GPS – global positioning system.
* $p < .05$ significant decline from baseline

Task Variables Affecting Recall

Two approaches were taken to better understand task variables that affect recall. In the first approach, six non-FBCB2 users completed the skill-retention measurement. These raters had no training or experience on FBCB2. Their instructions were to choose what appeared to them to be the most logical response to each question. As the raters had no prior training on FBCB2, the purpose for collecting these ratings was to determine whether the system's visual prompts or cues aided participants in answering the assessment questions.

The second approach was a *post hoc* characterization of the best and worst recalled questions. By examining a group of questions that were recalled well, it might be possible to identify those common characteristics facilitating recall. Similarly, for poorly recalled items, it might be possible to identify those common characteristics impairing recall.

For these analyses, the question was the unit of analysis. In other words, each step of each multi-step procedure was examined independently. For both approaches it was first necessary to rank order the questions by the percent of the Soldier sample able to recall the correct choice. The recall percentages were based on only the Soldiers who answered each question correctly at baseline. There was a total of 117 individual questions.

Non-FBCB2 User Performance. For the first approach, in which non-FBCB2 users served as raters, we examined both the number of correct raters and the total number of unique pairs of agreeing raters. The reason for counting pairs of agreeing raters was that there was often more than one correct response for each question. By measuring the number of agreeing raters, it was possible to index the degree to which the system pointed raters to the same response as opposed to just any correct response.

Two raters agreeing on a response constituted one pair of agreeing raters. Three raters that agreed on a single response constituted three unique pairs of agreeing raters and so on. When all six raters agreed on the same response, there was a total of 15 unique pairs of agreeing raters. The total number of agreeing pairs was summed for all correct and incorrect responses for each question. Thus, each question had two values from the raters, the number of pairs that agreed on correct responses and the number of pairs that agreed on incorrect responses.

Figure 2 summarizes the number of agreeing pairs of non-FBCB2 raters for each item. The items are ranked from high to low based on the percent of Soldiers who correctly recalled each item. The two lines indicate agreement on correct responses and agreement on incorrect responses. As can be seen, the non-FBCB2 raters had progressively less agreement on the correct responses and more agreement on the incorrect responses as Soldier performance decreased (agreement on incorrect responses was graphed using negative values so that the two lines would not overlap). The correlation between the number of agreeing pairs of non-FBCB2 raters and Soldier performance was $r = .52$ for correct responses and $r = -.30$ for incorrect responses. The correlation between the number of correct raters and the recall rate for each question was $r = .56$. Thus, one interpretation of these data is that the “intuitiveness” of the system can account for about 27% - 31% of the variance in retention. Conversely, about 9% of retention errors can be accounted for by system mis-cueing. Additional analyses of the non-FBCB2 user rater data can be found in Appendix E.

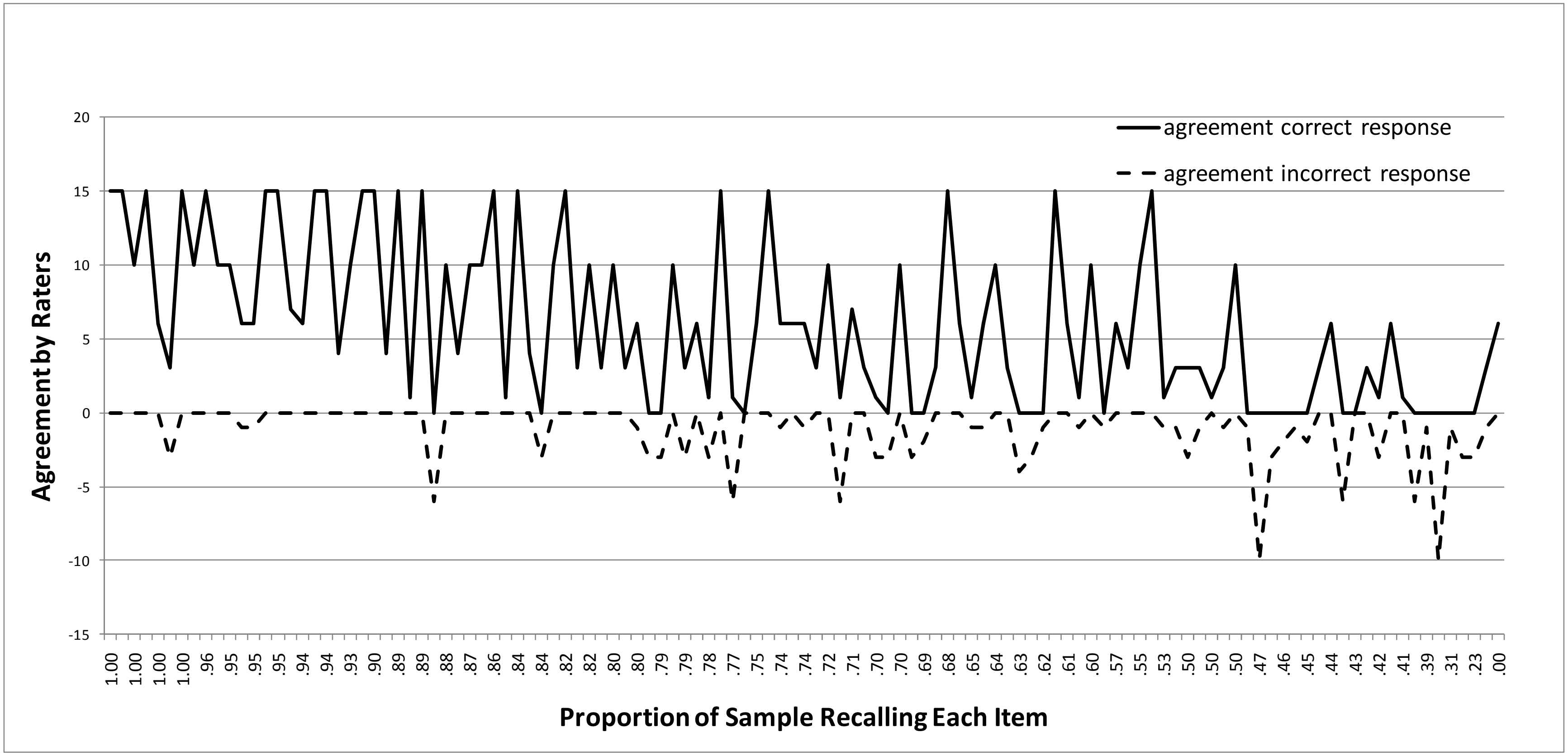


Figure 2. Agreement by non-FBCB2 users on correct and incorrect responses by the percent of Soldiers who recalled each item 8 weeks later. Agreement by raters is the number of unique pairs of the six raters agreeing on correct/incorrect responses for each question. Agreement on incorrect responses was graphed using negative values so that the two plots would not overlap. There are a total of 117 individual questions.

Characteristics of questions with good and poor recall. For a *post hoc* explanation of the results, we examined the best and worst recalled questions by the Soldier sample. Once again, the Soldier sample used was the group having perfect performance at baseline on each question (operations screen question or single step of a multi-step procedure). The group of good recall questions was comprised of those for which the recall rate was 90% or higher (23 questions). The group of poor recall questions was comprised of those for which the recall rate was 50% or lower (27 questions). These two breakpoints were chosen because they seemed reasonable from a training standpoint and because they included about 20% of questions each. It is important to note that this *post hoc* description is subjective in nature and is only meant to provide a possible explanation for differences in recall rate

Table 14
Characteristics of Questions with Good and Poor Recall

Characteristics	% of questions
Good recall	
Action in step is logically linked to the overall procedure or question	52%
Verification steps: e.g., “are you sure” or “ok”	39%
Item Completion: e.g., “close” or “apply”	13%
Poor recall	
Vague cues that don’t clearly indicate what needs to be done	44%
Misleading cues point to the wrong choice	33%
Forgot to perform some part of the procedure or repeated a step	22%

Questions with good recall could be divided into three general categories (Table 14). The largest category was comprised of questions that were logically linked to the procedure or question. An example of this category is the step in which Soldiers chose “clear logs and queues” from the start menu for the procedure *clear logs and queues.* Another example is selecting “SPOT report” from the list of report types when sending a SPOT report. Another example was completing a subcomponent of a procedure such as saving a message after creating it by selecting the “save” button or entering the details of a report in the appropriate data fields. Soldiers also did well when the system prompted them to verify an action such as going offline and on steps that completed items such as “close” or “apply.”

There were also three categories of questions that were poorly recalled. The largest category was comprised of questions for which cues in the system did not indicate subsequent steps. Several questions in this category resulted from confusion between the F5 Status, F6 Admin, F7 Apps, and Start buttons. These four buttons access a range of administrative, troubleshooting, and miscellaneous functions. For example, for *create a periodic reminder*, the correct choice was F7 Apps, but a common error was F6 Admin. When asked to troubleshoot a malfunctioning GPS, the correct choice was F5 Status but frequent errors were F6 Admin and Start.

In other cases, the availability of options was dependent on an action that was not clearly cued. For example, when attaching an overlay to an operations order, the “attachments…” button was grayed out. To activate it, the user had to select the “Order Thread” tab, but there was nothing to indicate that selecting that tab activated the “attachments…” button. When changing from CADRG map to satellite imagery, the “Imagery” checkbox was grayed out until the CADRG box was unchecked. Again, there was no prompt to indicate that unchecking the CADRG box would cause the satellite imagery box to become available.

Cues could also be misleading. For example, when manually placing their vehicle icon on the map, many Soldiers chose “F1 Map.” The correct choice was “F6 Admin.” When asked to use the circular line of sight button, many Soldiers chose the “LOS” (line of sight) button when the correct choice was “F7 Apps.” When asked to enter the MEDEVAC call-sign and voice net frequency, many Soldiers chose “F3 Combat Messages” where they see the call-sign and voice net frequency displayed on the MEDEVAC message. To change the voice net frequency and call-sign, they had to choose “F6 Admin.”

The final category of errors was forgetting a step or repeating a step that had previously been completed. For example when asked to create, save, and send a message, some Soldiers created and sent the message and then attempted to close before saving it. In another example, Soldiers attempted to add an icon to an overlay a second time.

Individual Difference Variables Affecting Recall

Individual differences also accounted for differences in recall and performance. As shown in Table 15, self ratings of FBCB2 proficiency predict performance on both baseline and retention measures. Positive correlations indicate that higher self-ratings of proficiency were related to higher overall performance on the measure. Self ratings accounted for 18% of the variability in baseline performance and 14% of the variability in retention performance.

Table 15
Correlations Between Individual Knowledge and Experience and Performance

	Baseline performance			Recall performance		
Individual measures	*r*	*p*	n	*r*	*p*	n
Self-rating of FBCB2 proficiency	.42	<.001	213	.38	<.001	62
Computer experience	.40	<.001	210	.48	<.001	62

The measure of computer experience (total of all types of computer experience) was also positively related to performance as seen in Table 15. More computer experience predicted better performance on the baseline and retention measure. Computer experience accounted for 23% of the variance in performance on the recall measure and 16% of the variance on the baseline score.

Table 16
FBCB2 Use While Deployed and Test Performance

	Baseline performance			Recall performance		
Individual measures	*M*	*SEM*	n	*M*	*SEM*	n
Used FBCB2 while deployed	.70	.02	40	.62	.15	12
Did not use FBCB2 while deployed	.71	.01	156	.62	.18	49

Using FBCB2 while deployed did not impact scores. When participants who used this system while deployed were compared to those who did not, *t*-test comparisons revealed no differences at baseline or recall (see Table 16). Furthermore, the number of months using FBCB2 while deployed did not significantly correlate with overall performance on either the baseline or retention measures.

Discussion

Digital Skill Retention

The first goal of this research was to examine skill retention following the standard Army operator course. Anecdotal reports indicating that digital skills are perishable are supported by the current research findings, but this depends heavily on both the way that skill retention is measured and on the procedure in question. Fewer than one third of Soldiers with perfect baseline performance were still able to perform 10 of the 12 multi-step procedures 8 weeks later. Fewer than half of these participants could perform 7 of the 14 operations screen questions 8 weeks later. These findings indicate that two months following training, there is substantial skill decay (i.e., forgetting) of the items if the criterion is perfect performance.

It is important to keep in mind that this retention measure was challenging. Soldiers had to provide the correct response on their first attempt. For multi-step procedures, this demand was even more challenging because Soldiers had to perform a whole sequence of steps correctly on the first attempt after an 8 week retention interval. When using the operational system, Soldiers would probably be able to overcome some decreases in proficiency levels by using trial and error to figure out how to perform various procedures.

Using the percent of steps correct, demonstrates that Soldiers do recall much of what they originally knew. Soldiers who performed perfectly at baseline recalled 70% or more of the steps for 10 of the 12 multi-step procedures. This pattern is consistent with a recent examination of skill retention for Soldiers after graduating from Basic Combat Training (Cobb, James, Graves, & Wampler, 2009). In that research project, when the criterion was performance of the whole procedure correctly, overall proficiency was worse than when the percent of steps performed correctly was the measure at both time points.

The examination of retention in the whole sample (i.e., regardless of whether the baseline response was correct) characterizes the overall level of skill decay that would be expected in a Soldier sample eight weeks following training. Overall performance on the multi-step procedures indicated a significant change from 72% to 65% correct. More skill decay was evident for the operations screen questions. On those questions, the typical Soldier performed about 72% of the 14 questions correct at baseline and about 52% correct at recall, a statistically significant decline of 20%.

Comparing the skill retention of Soldiers trained in traditional environments to those trained in dL environments revealed few differences. The findings for the overall results of the multi-step procedures indicate a significant effect of instructional environment but no interaction between that factor and time. As mentioned above, the main effect seems to reflect a slight sampling bias in the Soldiers in the dL sample who completed the retention measure. The lack of an interaction indicates that skill decay occurred at an equal rate in both instructional environments. That is, the better performance on the retention measure of the Soldiers in the dL classes reflects higher proficiency levels at baseline (i.e., higher skill acquisition levels) rather than better skill retention. The overall analysis of retention of on the operations screen questions found no differences in performance across the two instructional environments.

Effectiveness of Training FBCB2 Operators in a dL Environment

The second purpose of this research was to compare the training of a digital system in two different instructional environments: dL and traditional. Despite some demographic differences, student performance from the two instructional environments was comparable. Overall, the findings indicate that the instructional environment only affected baseline performance on the operations screen questions with the traditional students performing better than the dL students. It is unclear whether the source of this difference had to do with the instructional emphasis or the different backgrounds of the two samples.

Whatever the reason, neither the analysis of the multi-step procedures nor the overall analysis of all the questions indicated any differences in baseline performance between dL and traditional students. Thus, the effect of instructional environment was subtle and limited to operations screen questions. Finally, as mentioned in the prior section, the instructional environment did not affect skill retention in any of the overall analyses.

Factors Contributing to Skill Decay

As mentioned in the introduction, there are three types of variables that are known to contribute to forgetting: training variables, task variables, and individual variables. Although the two training groups received comparable training, the results demonstrated that both task and individual variables had an impact on skill retention.

System cues, a type of task variable, had the largest effect on recall. The data from both approaches (examining non-FBCB2 user performance and analyzing the characteristics of questions) supported this conclusion.

The data from non-FBCB2 users showed that 27% - 31% of the variance in recall on our measure could be accounted for by the general cues present in the system depending on whether agreement among non-FBCB2 raters or simply correct responses are used as the basis for comparison. Overall, the findings from the non-FBCB2 users indicated that system mis-cueing accounted for little variance in recall on our measure. Rater agreement on incorrect responses only accounted for about 10% of the variance in skill decay.

Further, the findings suggest that vague cues also contributed to poor performance on the retention task. The lack of clear cues for subsequent steps was found to contribute substantially to poor performance in prior research on the retention of Soldier skills (Rose, Czarnolewski, Gragg, Austin, & Ford, 1985). As FBCB2 software continues to evolve, eliminating vague cues will certainly help to improve skill retention.

Regarding individual variables, both self-ratings of proficiency on FBCB2 and prior computer experience predicted performance on the measure at both time points. Interestingly, those who used FBCB2 on a deployment did not perform better than those who did not nor did the number of months using it predict performance. Based on these findings, it is the level of proficiency rather than experience that predicts performance.

Recommendations for Training.

Until the system is re-designed, instructors will need to provide Soldiers with better ways to remember some of these procedures. Below are some suggestions derived from the research findings that might be of use to training developers and instructors for both dL and traditional instructional environments.

Develop training to help Soldiers distinguish between the “Start”, “F7 Apps”, “F6 Admin”, and “F5 Status” buttons (all on the operations screen). Confusion about the functions accessed through these buttons appeared to be a cause of poor performance on a number of steps. Calling attention to the distinctions between functions under these buttons or developing exercises and job aides to help Soldiers remember the functions would be helpful. For example, instructors could develop mnemonics, or have Soldiers develop their own mnemonics, to help remember function access.

Call attention to places where system cues are especially vague or inconsistent. For example, there are many different ways to add objects and icons to the map and to overlays. In the overlay toolbox, on the “group setup” tab, the “add icon” button adds a selected object to the overlay object group. On the “object” tab, the “add” button allows the user to place a selected object on the map, and in other places, icons are placed on the map by use of a “map” button. It is easy to see how over time confusion arises over the functions of these buttons.

Call attention to FBCB2 conventions that differ from those found in Windows. For example, when creating a folder in Windows, the user first tells the system to create a new folder and then names it. In FBCB2, the user first enters a name and then tells the system to create a new folder on the “manage” tab under “F4 Messages.” Calling attention to these inconsistencies will help Soldiers to encode the differences between Windows and FBCB2. Additionally, research on training techniques to overcome some of the interface deficiencies would benefit trainers and training developers.

Conclusions

In conclusion, it is important for unit leaders to understand that Soldiers show significant forgetting of FBCB2 skills as early as 8 weeks following classroom instruction. Though research on the time needed for refresher training was not specifically examined in the present research, the fact that for most of the multi-step procedures, Soldiers forgot less than a third of the steps at 8 weeks, suggests that Soldiers could be retrained relatively quickly. Prior research on skill retraining suggests that it typically takes about half as much time to retrain Soldiers as it does to train the first time (Wisher, Sabol, Ellis, & Ellis, 1999). Anecdotal comments by Soldiers in the present research indicated that it may even require less time than that. Typically, Soldiers felt that they could refresh their skills within a couple of hours by just "playing" with the system. Future empirical research is needed to determine how much refresher training is needed to restore proficiency.

Finally, the findings of this report indicate that dL instruction as examined in the present research is an effective way to train FBCB2 skills. Both baseline performance immediately following training and the retention of skills over the 8 week interval was comparable for both training environments. While this does not mean that all dL instruction would be expected to be as effective as all traditional training, it does indicate that it is possible to deliver effective digital system training in either mode of instruction.

References

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention , and transfer of human motor skills. *Psychological Bulletin*, 101(1), 41-74.

Bink, M. L., Wampler, R. L., Goodwin, G. A., & Dyer, J. L. (2009). *Combat veterans' use of Force XXI Battle Command Brigade and Below (FBCB2)* (ARI Research Report 1888). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADB347437)

Bonk, C. J., & Graham, C. R. (2006). *The handbook of blended learning: Global perspectives, local designs.* San Francisco, CA: Pfeiffer.

Clark, J. E. (2005). *Solving a command and control system education and training dilemma for the modular force (A white paper).* Washington, DC: Army Joint Support Team.

Deatz, R. C., & Campbell, C. H. (2001). *Application of cognitive principles in distributed computer-based training* (ARI Research Product 2001-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA392929)

Dyer, J. L., Fober, G. W., Wampler, R., Blankenbeckler, N., Dlubac, M., & Centric, J. (2000). *Observations and assessments of Land Warrior Training*. Ft. Benning, GA: U.S. Army Research Institute for the Behavioral and Social Science, Infantry Forces Research Unit. (Limited Distribution)

Dyer, J. L., Singh, H., & Clark, T. L. (2005). *Computer-based approaches for training interactive digital map displays* (ARI Research Report 1842). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA440171)

Goodwin, G. A. (2006). *The training, retention, and assessment of digital skills: A review and integration of the literature* (ARI Research report 1864). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA470707)

Goodwin, G. A., Leibrecht, B. C., Wampler, R. L., Livingston, S. C. & Dyer, J. L. (2007). *Retention of selected FBCB2 operating skills among Infantry Captains Career Course (ICCC) students.* (ARI Research Report 1872). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA470741)

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. *Human Factors, 25*(2), 199-213.

Leibrecht, B. C., Lockaby, K. J., & Meliza, L. L. (2003a). *Exploiting FBCB2 capabilities through realistic feedback* (ARI Research Report 1810). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA415999)

Leibrecht, B. C., Lockaby, K. J., & Meliza, L. L. (2003b). *A practical guide for exploiting FBCB2 capabilities* (ARI Research Product 2003-05). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA415997)

Leibrecht, B. C., Lockaby, K. J., Perrault, A. M., & Meliza, L. L. (2004a). *Digital proficiency levels for the brigade and battalion battle staff* (ARI Research Report 1826). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA425934)

Leibrecht, B. C., Lockaby, K. J., Perrault, A. M., & Meliza, L. L. (2004b). *Measuring digital battle staff proficiency in current and future forces* (ARI Research Report 1825). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. B301959)

Leibrecht, B. C., Lockaby, K. J., Perrault, A. M., Strauss, C. P., & Meliza, L. L. (2006). *Tailored exercise planning and feedback for digitized units* (ARI Research Report 1858). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (Restricted Distribution)

Leibrecht, B. C., Wampler, R. L., Goodwin, G. A., & Dyer, J. L. (2007). *Techniques and practices in the training of digital operator skills.* (ARI Research Report 1878). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA474556)

Sanders, W. R. (2001). *Cognitive psychology principles for digital systems training* (ARI Research Report 1773). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA 391035)

Rao, S. (2001). InterForm [Software]. Seattle, WA: Raosoft.

Rose, A. M., Czarnolewski, M. Y., Gragg, F. E., Austin, S. H., & Ford, P. (1985). *Acquisition and retention of Soldiering skills* (ARI Technical Report 671). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA160336)

Tucker, J. S., McGilvray, D. H., Leibrecht, C. B., Strauss, C., Perrault, A., & Gesselman, A. N. (2009). *Training digital skills in distributed classroom environments: A blended learning approach* (ARI Research Report No. 1893). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA495731)

Wisher, R. A., Sabol, M. A., Ellis, J., & Ellis, K. (1999). *Staying sharp: Retention of military knowledge and skills* (ARI Special Report 39). Alexandria, VA: U.S. Army Research Institute for the Social and Behavioral Sciences. (DTIC No. ADA366825)

Acronyms

ABCS	Army Battle Command System
AFATDS	Advanced Field Artillery Tactical Data System
ASAS	All Source Analysis System
BCS3	Battle Command Sustainment Support System
CPOF	Command Post of the Future
GPS	Global Positioning System
FBCB2	Force XXI Battle Command Brigade and Below
ICCC	Infantry Captains' Career Course
MCS	Maneuver Control System
MGRS	Military grid reference system
MOS	Military occupational specialty
OPORD	Operation Order
SA	Situation Awareness
SINCGARS	Single Channel Ground and Airborne Radio System
SPOT	Not an acronym for anything. Same as size, activity, location terrain report

Appendix A

Demographic Data

Question	Responses	dL Baseline Only Count	dL Baseline Only %	dL Baseline+ Recall Count	dL Baseline+ Recall %	Traditional Baseline Only Count	Traditional Baseline Only %	Traditional Baseline+ Recall Count	Traditional Baseline+ Recall %
What is your current grade/rank?	E3 (PFC)	6	5.9%	1	3.1%	14	28.6%	10	33.3%
	E4 (SPC/CPL)	28	27.5%	9	28.1%	16	32.7%	8	26.7%
	E5 (SGT)	30	29.4%	9	28.1%	6	12.2%	6	20.0%
	E6 (SSG)	24	23.5%	8	25.0%	6	12.2%	2	6.7%
	E7 (SFC)	7	6.9%	0	.0%	2	4.1%	2	6.7%
	E8 (MSG / 1SG)	4	3.9%	3	9.4%	0	.0%	0	.0%
	E9 (SGM / CSM)	0	.0%	0	.0%	0	.0%	0	.0%
	01/02 (LT)	2	2.0%	1	3.1%	5	10.2%	2	6.7%
	03 (CPT)	1	1.0%	1	3.1%	0	.0%	0	.0%
	04 (MAJ)	0	.0%	0	.0%	0	.0%	0	.0%
	O5 (LTC)	0	.0%	0	.0%	0	.0%	0	.0%
If you were to deploy tomorrow, how likely is it that you would use FBCB2?	Very unlikely	10	9.8%	3	9.4%	10	20.4%	3	10.0%
	Somewhat unlikely	17	16.7%	4	12.5%	5	10.2%	2	6.7%
	Somewhat Likely	44	43.1%	18	56.2%	18	36.7%	11	36.7%
	Very Likely	31	30.4%	7	21.9%	16	32.7%	14	46.7%

		dL				Trad			
		Baseline Only		Baseline+ Recall		Baseline Only		Baseline+ Recall	
Question	Responses	Count	%	Count	%	Count	%	Count	%
If you were to deploy tomorrow, in what duty position/role would you most likely use FBCB2?	Unknown	30	29.4%	7	21.9%	18	36.7%	10	33.3%
	Primary Operator for a Leader	6	5.9%	4	12.5%	7	14.3%	5	16.7%
	Section LDR /Squad LDR	26	25.5%	6	18.8%	8	16.3%	9	30.0%
	Vehicle CDR (other than LDR/CDR)	7	6.9%	6	18.8%	4	8.2%	2	6.7%
	PLT LDR / PSG	4	3.9%	1	3.1%	8	16.3%	3	10.0%
	CO HQs or support element	13	12.7%	3	9.4%	3	6.1%	0	.0%
	CO CDR / Troop CDR / 1SG	0	.0%	1	3.1%	0	.0%	0	.0%
	Staff Officer/NCO in BN or BDE TOC	16	15.7%	4	12.5%	1	2.0%	1	3.3%
Overall how would you rate your proficiency on FBCB2 ?	Never Used	15	14.7%	1	3.1%	1	2.0%	0	.0%
	Basic	47	46.1%	13	40.6%	5	10.2%	5	16.7%
	Medium	36	35.3%	15	46.9%	37	75.5%	20	66.7%
	High	4	3.9%	3	9.4%	6	12.2%	5	16.7%
Have you ever used FBCB2 while deployed on a combat tour ?	Yes	16	18.4%	3	9.7%	12	25.0%	9	30.0%
	No	71	81.6%	28	90.3%	36	75.0%	21	70.0%
Overall how would you rate your proficiency on ASAS ?	Never Used	72	70.6%	26	81.2%	35	71.4%	27	90.0%
	Basic	20	19.6%	5	15.6%	5	10.2%	3	10.0%
	Medium	9	8.8%	1	3.1%	9	18.4%	0	.0%
	High	1	1.0%	0	.0%	0	.0%	0	.0%

		dL				Trad			
		Baseline Only		Baseline+ Recall		Baseline Only		Baseline+ Recall	
Question	Responses	Count	%	Count	%	Count	%	Count	%
Have you ever used ASAS while deployed on a combat tour ?	Yes	0	.0%	0	.0%	0	.0%	0	.0%
	No	32	100.0%	6	100.0%	16	100.0%	5	100.0%
Overall how would you rate your proficiency on AFATDS ?	Never Used	84	82.4%	29	90.6%	42	85.7%	25	83.3%
	Basic	13	12.7%	2	6.2%	4	8.2%	3	10.0%
	Medium	2	2.0%	0	.0%	3	6.1%	2	6.7%
	High	3	2.9%	1	3.1%	0	.0%	0	.0%
Have you ever used AFATDS while deployed on a combat tour ?	Yes	0	.0%	0	.0%	1	14.3%	0	.0%
	No	18	100.0%	3	100.0%	6	85.7%	6	100.0%
Overall how would you rate your proficiency on MCS ?	Never Used	74	72.5%	23	71.9%	36	73.5%	24	80.0%
	Basic	19	18.6%	6	18.8%	11	22.4%	4	13.3%
	Medium	7	6.9%	2	6.2%	2	4.1%	2	6.7%
	High	2	2.0%	1	3.1%	0	.0%	0	.0%
Have you ever used MCS while deployed on a combat tour ?	Yes	3	10.3%	5	55.6%	0	.0%	1	14.3%
	No	26	89.7%	4	44.4%	13	100.0%	6	85.7%
Overall how would you rate your proficiency on CPOF ?	Never Used	87	85.3%	29	90.6%	43	87.8%	27	90.0%
	Basic	10	9.8%	2	6.2%	2	4.1%	1	3.3%
	Medium	4	3.9%	0	.0%	4	8.2%	2	6.7%
	High	1	1.0%	1	3.1%	0	.0%	0	.0%
Have you ever used CPOF while deployed on a combat tour ?	Yes	1	6.7%	1	33.3%	1	20.0%	0	.0%
	No	14	93.3%	2	66.7%	4	80.0%	3	100.0%

Appendix B

Knowledge Test Questions

Knowledge question baseline comparison. The percent of knowledge questions answered correctly at baseline was computed for each Soldier and this overall knowledge question measure was analyzed in a two factor, number of tests by instructional environment, ANOVA. There was only a significant interaction effect, $F(1,209) = 10.9$, $p < .01$. As with the multi-step procedures, the dL participants that returned for the retention measure scored higher at baseline (89% correct) than those that didn't (79% correct), $F(1,209) = 5.8$, $p < .05$, whereas the traditional students who returned scored lower (75% correct) than those that didn't (87% correct), $F(1,209) = 5.2$, $p < .05$.

A comparison of baseline performance on each knowledge question by the dL and traditional students was also performed. These were multiple choice questions and so the performance measure was the percent of the sample correctly answering each question. These data were analyzed with chi-square tests. The chi-square results indicated no significant differences between the two instructional environments for any of the knowledge questions (Table B-1).

Table B-1
Baseline Results for Knowledge Questions

Knowledge Questions - Baseline	Average % Correct Rank ordered by dL scores	
	dL (n = 133)	Traditional (n = 80)
A flashing number in the area indicated above means… *A new message has been received*	96%	91%
The area indicated above serves what function? *To pan the view around the map*	93%	93%
Why is the send button on the SPOT report indicated above grayed out? *The location field has not been completed*	85%	80%
Clicking on the red box indicated above brings up which menu? *Status*	74%	79%
What are the fields indicated above used for? *Choosing operational symbols for an overlay*	60%	69%

Knowledge questions retention comparison. Performance on the knowledge questions by the traditional and dL classes across the retention interval was compared in a two factor, time by instructional environment, repeated measures ANOVA. The percent of the five questions answered correctly was used as an overall performance measure. The ANOVA revealed only a significant effect of time (Wilks' $\Lambda = .90$, $F\ (1, 52) = 6.05$, $p < .05$). Table B-2 shows both the baseline and retention average percentages for the traditional and dL participants for this subsample. Table B-2 also shows the overall average percent decrease in performance for these questions across both groups.

Table B-2
Retention Results for Knowledge Questions

Knowledge Questions (*Average of 5 Items*)	dL		Traditional		Average % Decrease
	Baseline	Retention	Baseline	Retention	
Retention sample	90%	79%	75%	68%	9%

To better understand the knowledge questions for which Soldiers' performance decreased the most over time, Table B-3 presents the average baseline and retention scores for each question. Wilcoxon tests revealed that the only significant change in performance was for the dL classes on the *Clicking on the red box to bring up the Status menu* question.

Table B-3

Percent Correct Scores for All Knowledge Questions – Retention Sample Only

Knowledge Questions	dL *Rank ordered by Retention scores*		Traditional	
	Baseline	Retention	Baseline	Retention
A flashing number in the area indicated above means… *A new message has been received*	100%	100%	87%	81%
The area indicated above serves what function? *To pan the view around the map*	91%	96%	87%	77%
Why is the send button on the SPOT report indicated above grayed out? *The location field has not been completed*	94%	87%	65%	77%
What are the fields indicated above used for? *Choosing operational symbols for an overlay*	69%	65%	68%	45%
Clicking on the red box indicated above brings up which menu? *Status*	94%[1]	48%[1]	71%	58%

Notes. [1] = Significant performance decrements for Soldiers in the dL classes.

Appendix C

Summary of Hands-on Portion of Skill Retention Measure

Single-step Operations Screen Questions

The first set of 14 questions all focused on the operations screen. The operations screen is the default screen of FBCB2. It includes the situation awareness display (map with icons), the function and Quicksend buttons on the right side, status indicators on the bottom and the flash, immediate, priority, and routine (FIPR – pronounced "fipper") queue on the top (see Figure C-1).

In all of the operations screen questions, users were asked to indicate the first step to perform a range of procedures. Operations screen questions, and all other questions, were written as multiple choice questions. Responses to each question corresponded to choices that were available on the screenshot. A letter was assigned to each response and was superimposed on the screenshot to help users unambiguously identify each response.

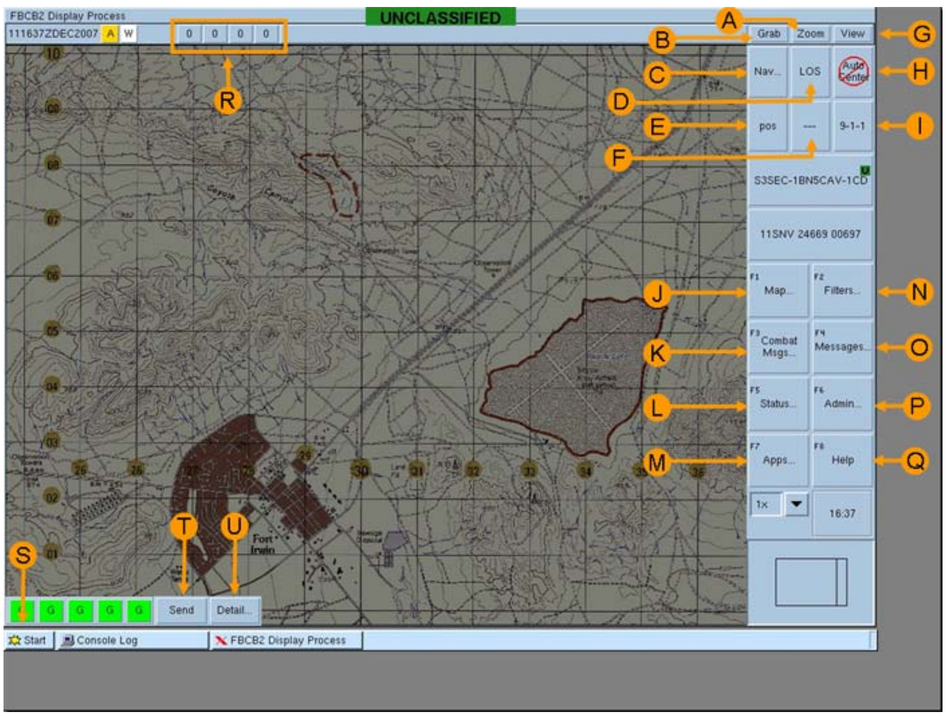


Figure C-1. A view of the operations screen. The situation awareness (SA) display is the map. The flash, immediate, priority, routine (FIPR) queue is indicated at R. The panel on the right includes the function keys (J – Q), quicksend buttons (E, F, & I), as well as various other map tools and location indicators. The status indicator is in the lower left (five green gumballs). The operations screen response options were the same.

The 16 operations screen questions are below. The correct choice is indicated in parentheses.

1. Your GPS is not operating at the optimum level and you need to troubleshoot the problem. The hardware has been checked and there were no problems. You need to check the software. What is the first step? (L – F5 Status) Note: the screenshot for this question was different than Figure C-1 in that the two "gumballs" on the top left corner of the screen were both red which would indicate trouble with the GPS.
2. You must manually place your icon/platform on the map. What is the first step? (P – F6 Admin)

3. You must enter the MEDEVAC call sign and voice net frequency. What is the first step? (P – F6 Admin)
4. You must create a periodic reminder. What is the first step? (M – F7 Apps)
5. You must create an address group. What is the first step? (O- F4 Messages)
6. A message has been created. Assign it to a quick send button. What is the first step? (F - --- or O- F4 Messages)
7. You must create and save a new position report. What is the first step? (O – F4 Messages)
8. You must activate the driver's display for a route. What is the first step? (C – Nav)
9. You must use the circular line of sight tool. What is the first step? (M – F7 Apps)
10. You must display an overlay message that was just sent to you. What is the first step? (R – 0001) Note: The screenshot for this question was slightly different from figure C-1 in that the FIPR queue read 0001 indicating that a new message had been received.
11. You must create and send an NBC 1 report. What is the first step? (K – F3 Combat Messages or O – F4 Messages)
12. You must create and send a Mayday report. What is the first step? (K – F3 Combat Messages, I – 9-1-1, or O – F4 Messages)
13. Your vehicle status has been updated. You must transmit your combat platform status/SITREP. What is the first step? (T – Send)
14. You must edit a location folder. What is the first step? (J – F1 Map)

Multiple-Step Full Procedure Questions

There were seven full procedure questions. In these questions, participants started with the operations screen and had to complete all steps of each procedure. What follows is a description of each of these procedures.

<u>Create a message folder (5 steps)</u>. Folders can be created to help organize messages in FBCB2. This procedure starts with the “F4 Messages” function and then the “manage” tab. From there the user types the name of the folder in a field and then clicks on the “create” button. A common error at recall was to click the “create” button before entering the name of the new folder. This may have been due to interference from the convention of common commercial operating systems in which the sequence is typically to create and then name the folder.

<u>Set default addressing for a SPOT report (8 steps)</u>. To save time during operations, FBCB2 allows users to pre-enter default addresses for the various types of reports. In this procedure, users are also told to set the default “precedence” and “acknowledge” settings for this type of report. Users set the “precedence” to “flash” (the highest precedence level) and “acknowledge” to “operator acknowledge.” The “acknowledge” setting would require the recipient to manually indicate when the message is received. To select recipients, users must search for recipients and manually add them to a list (see screenshot below).

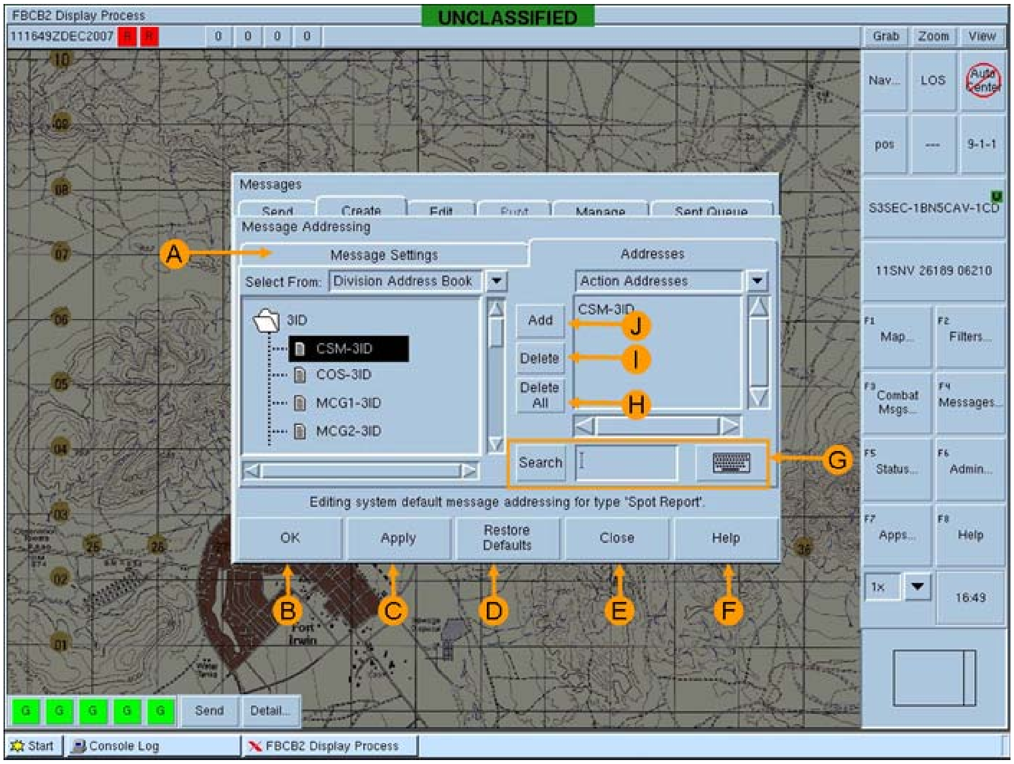


Figure C-2. A view of the default message addressing dialogue box. After using the search function to find the name of the recipient, the user highlights that name and then chooses "add" to add that name to the list. Precedence and acknowledge settings are done on the "message settings" tab.

Clear logs and queues (11 steps). This procedure is performed to enhance system performance that has degraded because there is too much stored data. As users go through this process, they have the option to selectively remove types of files. This is one of the more easily forgotten procedures because the first few steps are not well cued. The first step involves exiting the operations screen through the "F6 - Admin" function then the user goes "offline" and chooses the "start" button and then the "FBCB2" option. It is not until this point that the user sees the "clear logs and queues" option.

Create, save, and send a SPOT report (7 or 4 steps, depending on the path chosen). This was a procedure that branched depending on user input. Branching was based on whether the user chose "F3 Combat Msgs" or "F4 Msgs." The F4 choice requires users to choose "SPOT report" from a larger list, whereas the F3 choice takes them directly to the "combat messages" screen (see Figure C-3). This procedure requires users to create, save, and send one of the many pre-formatted message types found in FBCB2. The pre-formatted messages require the user to enter data into fields of the message. In the case of the SPOT report, this makes it possible for the data to automatically appear on the map of each recipient's system. For example, if the sender sees two enemy vehicles at some location and sends out a SPOT report, icons for the two enemy vehicles will appear at the correct location on every recipient's map.

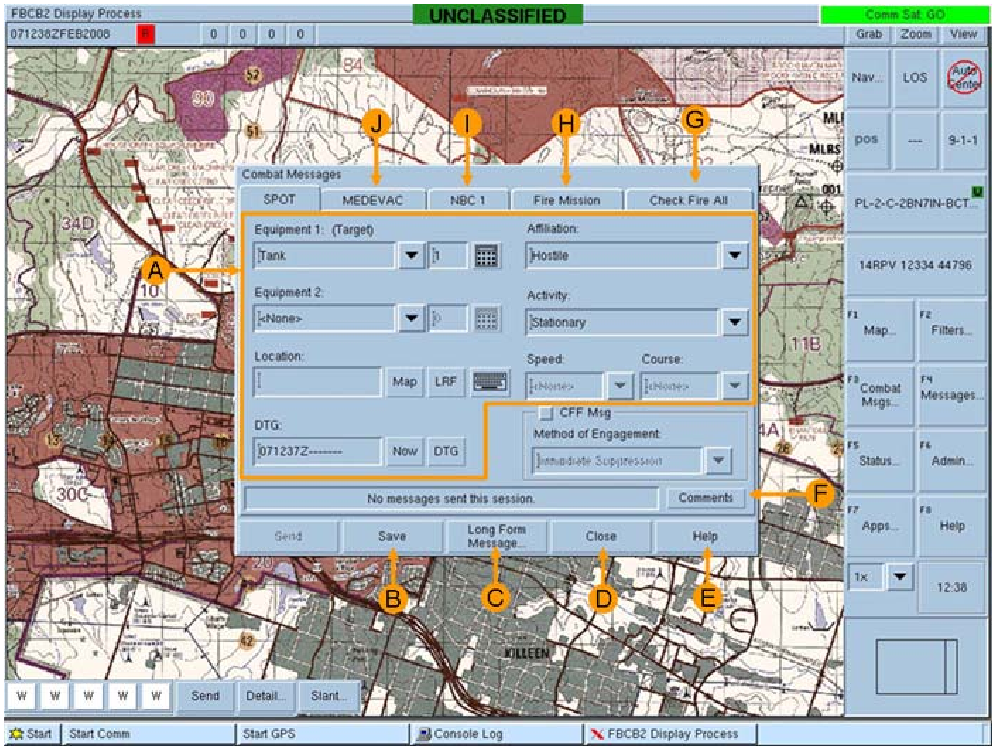


Figure C-3. The combat messages dialogue box showing the SPOT report screen. The various fields of the SPOT report are completed by using drop down menus. The location field can be completed manually or by clicking a location on the map. The date-time-group (DTG) field can be completed by clicking the "now" button or by manually entering a DTG.

<u>Create and send a route (11 Steps)</u>. One of the primary functions of FBCB2 is to aid in vehicle navigation. Creating a route in FBCB2 not only plots the route on the map, but also provides data on elevation changes and grade across the route. Routes can be sent as messages so that all recipients can follow or see the route. The route can be created by either clicking on the map to add waypoints or by entering the exact grid coordinates of each waypoint manually (Figure C-4). Once created, the route can be sent as a message using the same basic message addressing and sending process as for other messages and reports.

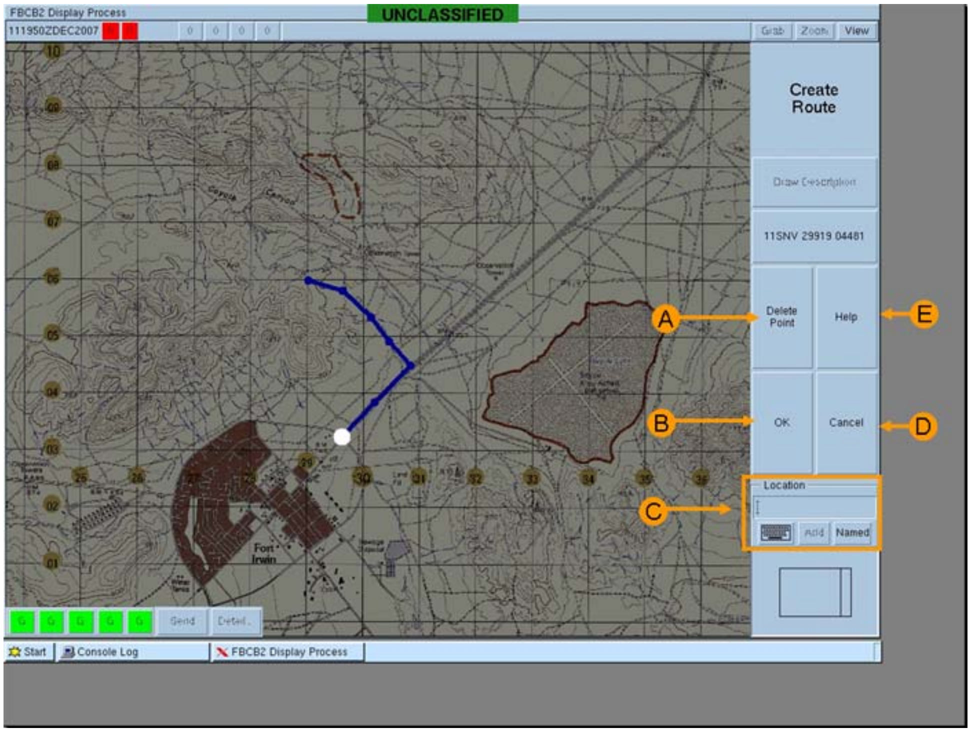


Figure C-4. Screenshot of the interface for creating a route. Clicking on the map places a waypoint at that location. Waypoints can also be added by entering the exact grid coordinates in the field at C.

<u>Set your screen to display all enemy units and only current, friendly units (6 or 7 steps)</u>. The map in FBCB2 can potentially become cluttered with many icons of friendly or enemy units. To help reduce the number of icons displayed, filters can be applied. This step begins with the "F2 Filters" button and is a relatively well-cued procedure that involves checking boxes to indicate what to display. Branching in this question is based on the sequence in which certain boxes are selected.

<u>Show specified vehicle on the display (6 steps).</u> In this procedure, users need to find a specific vehicle and center the view on that vehicle. Users first must select "F1 Map" which shows the map control dialogue box. From here they choose the "center" tab and then choose the "unit/platform" tab. From here they can choose the vehicle from a list or they can use a search function to find the vehicle. In this question, the vehicle they were to center on was displayed on the list.

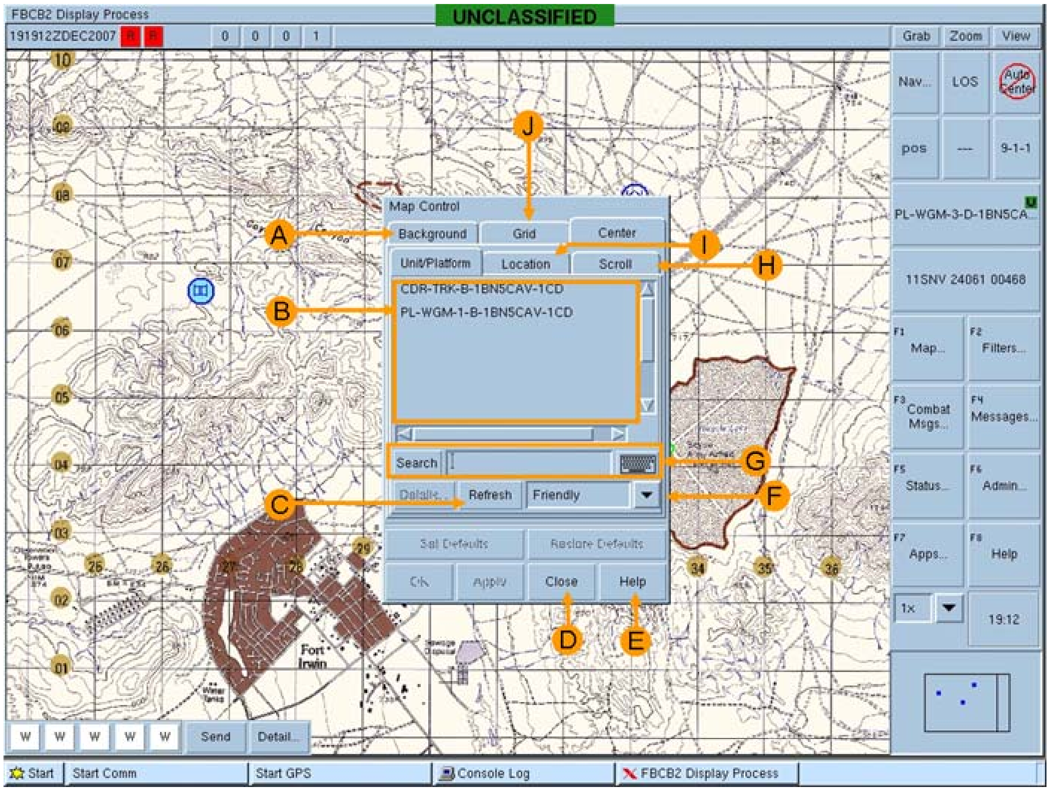


Figure C-5. View of the map control dialogue box. The “center” tab has three subtabs: “unit/platform” which is shown, “location”, and “scroll.” The “grid” and “background” tabs are for adding a grid of a selected scale or adding a

Multiple-Step Partial Procedure Questions

There were five multiple-step partial procedure questions. Partial procedure questions included a subset of the total steps. In all of these questions, Soldiers were told which steps of the procedure had already been completed prior to the first step shown. Soldiers then had to complete the remaining steps.

Create and save an overlay object group (11 or 12 Steps). An overlay object group is a folder in which a set of commonly used standard military symbols can be stored for easy access. Otherwise, symbols are found by searching from a catalogue of thousands of symbols. It should be noted that building overlays is not a procedure that junior enlisted Soldiers would often perform, but these are skills that are taught in the operator course. In this question, the overlay toolbox has been opened and Soldiers must finish the procedure from there. The procedure involves using the “group setup” tab to create and name the group and then the “object” tab to search for and select objects one by one (Figure C-6). Once an object is selected, the group setup tab is used to add each object to the group. Branching in this procedure was based on whether the user started by creating a new group or by selecting the object.

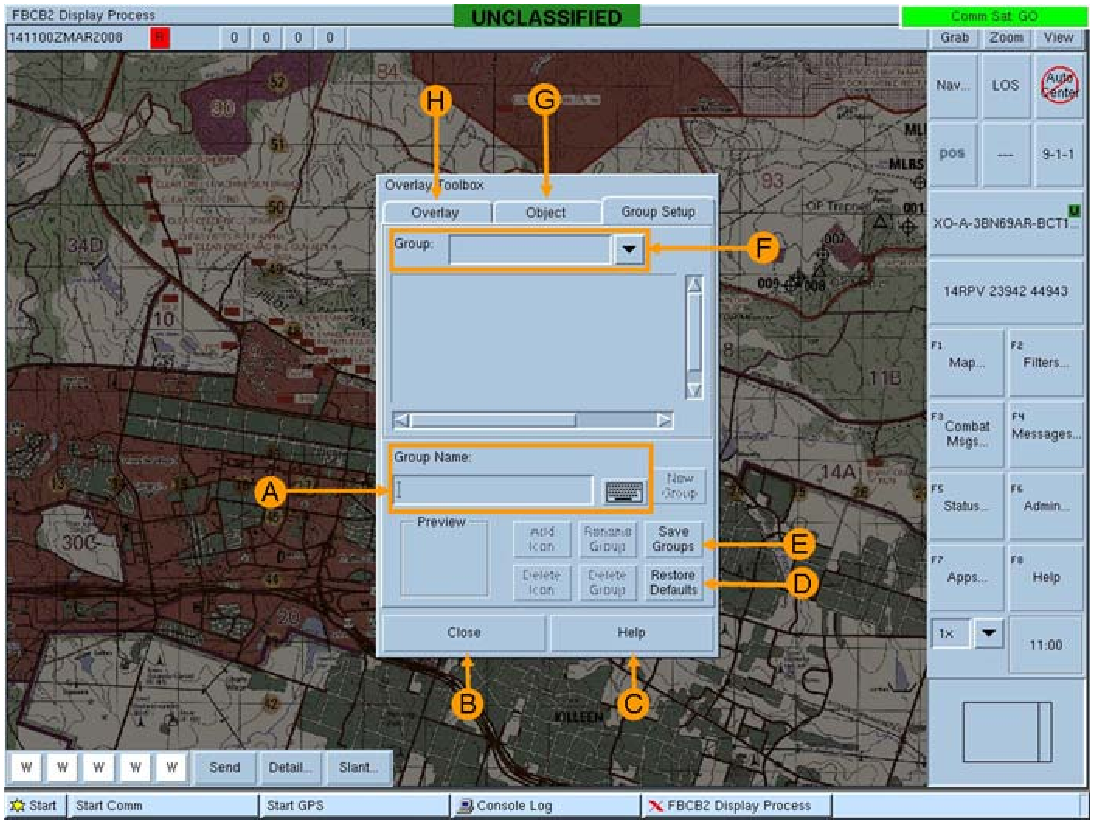


Figure C-6. The overlay toolbox. This dialogue box has three tabs: the “overlay” is for creating an overlay, the “object” tab is for selecting objects, and the “group setup” tab (shown) is for creating an overlay object group.

<u>Create and save a combat services support overlay (11 steps).</u> In this procedure, Soldiers had to create a combat service support overlay and indicate the location of a bulldozer. They were to leave the overlay displayed when finished. The “overlay toolbox” is displayed at the beginning of this procedure and Soldiers had to finish the procedure from there. This procedure is similar to the previous one in that the “overlay” tab is used to create the overlay and the “object” tab is used to add objects to that overlay. The difference is that objects are added to the overlay by clicking on the map. Branching in this procedure was determined by whether the Soldiers decided to create the overlay or select the object first.

<u>Attach an overlay to an Operations Order (OPORD) (6 Steps)</u>. In this procedure, the Field Order Management Tool has been opened from the F4 Messages function (Figure C-7). The Soldier sees the OPORD on a list on the “Field Orders” tab, but the “Attachments” button is grayed out. To add an attachment, the “Order Thread” tab must be selected which shows details for this OPORD and is virtually identical to the “Field Orders” tab. After selecting the OPORD from the list, the “Attachments” button become active and clicking it allows the Soldier to select the overlay to attach and then attach it. The “Attachments” button is a clear cue for this procedure, but it is not obvious that selecting the “Order Thread” tab and then selecting the OPORD should be necessary to activate this button. As described in the report, this was one of the most forgotten steps.

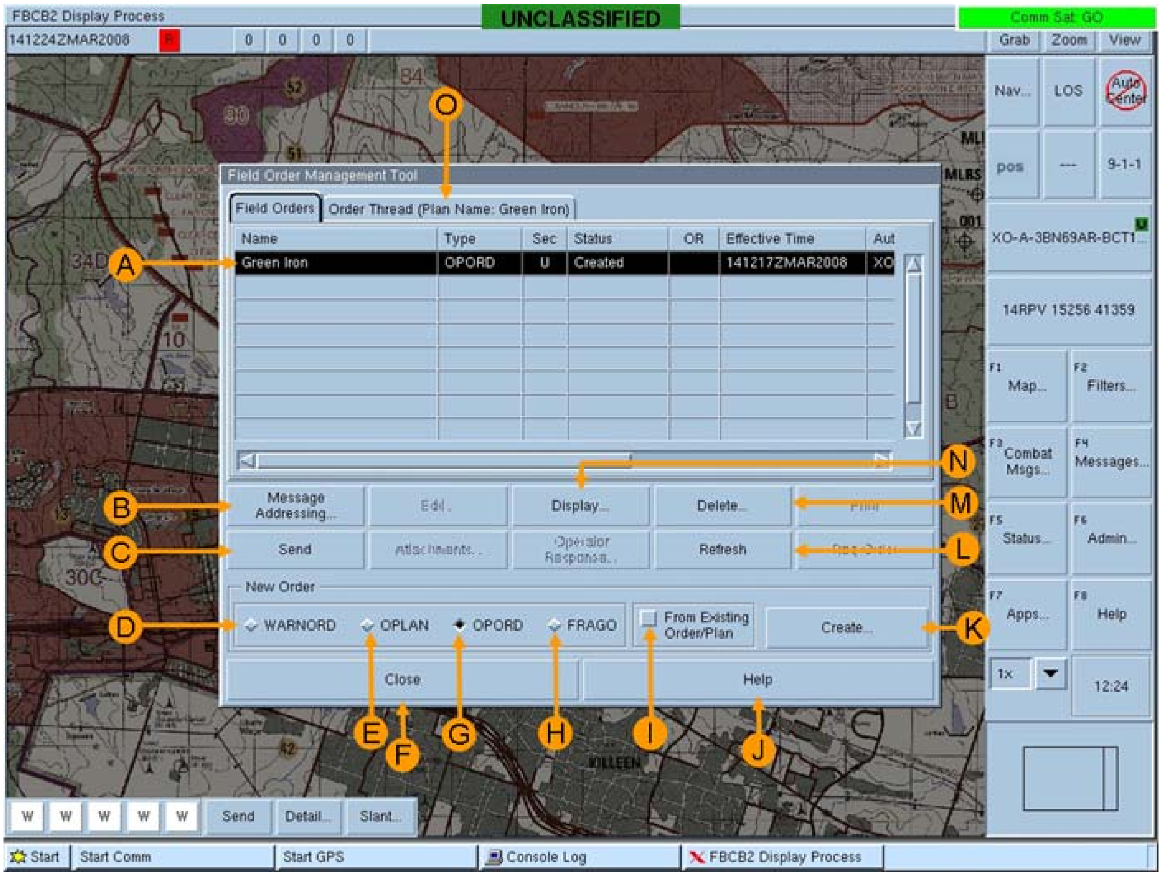


Figure C-7. The Field Order Management Tool has two tabs: Field Orders and Order Thread. Note that the "Attachments" button is grayed out. Selecting the "Order Thread" tab and then clicking on the order will activate this button.

Display a satellite image on the SA display (3 steps). This is a relatively short procedure. The Soldier starts with the "Map Control" dialogue box open (Figure C-8). This dialogue box is accessed by choosing the "F1 Map" function. The "Background" tab is in the foreground by default and there is an "imagery" checkbox available; however, it is grayed out when the "CADRG" box is checked. Both "CADRG" and "Imagery" cannot be selected at the same time. To make the "imagery" checkbox active, the "CADRG" checkbox must first be unchecked. This step was often forgotten, possibly because of interference from common commercial operating systems in which selection of one would automatically unselect the other.

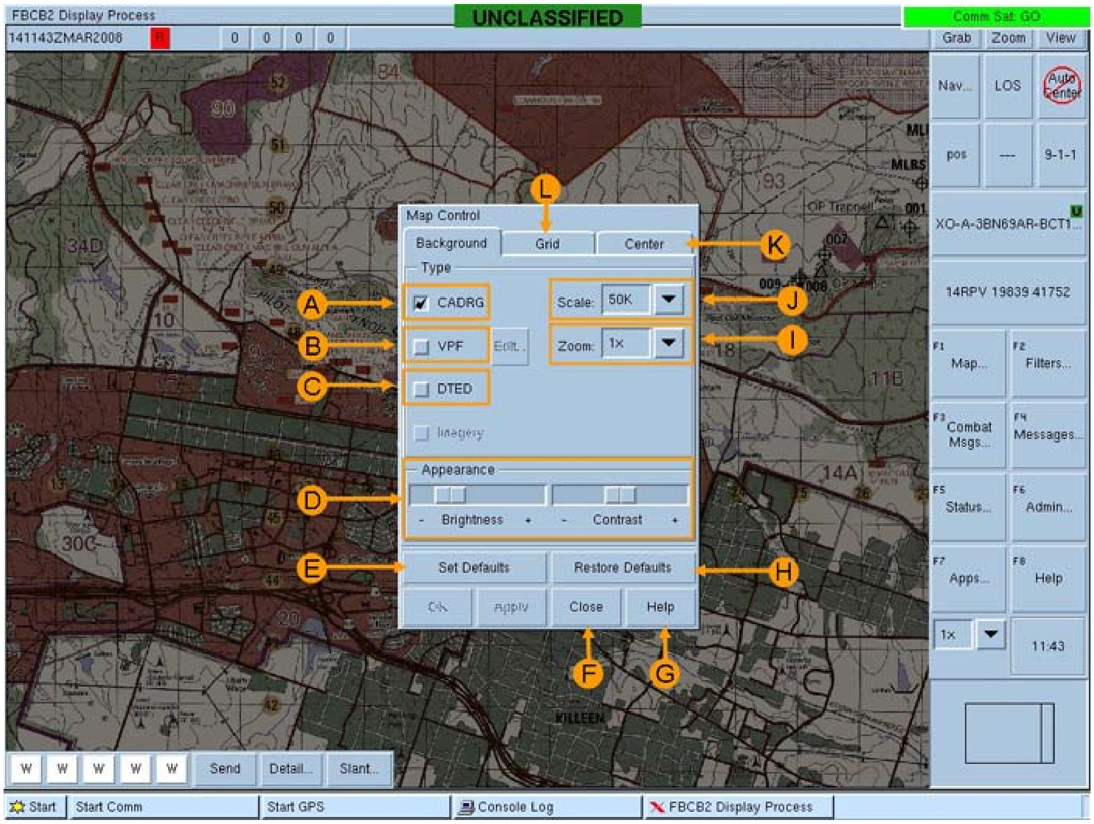


Figure C-8. The "Map Control" dialogue box accessed through the F1 Map function. Note the grayed out "Imagery" checkbox.

<u>Display gridlines on your map (4 Steps).</u> This procedure begins with the "Map Control" dialogue box open. Soldiers are instructed to add gridlines using the military grid reference system (MGRS) coordinates with 10 km spacing and green color. The user must choose the "Grid" tab and then has the option to set the "grid color" and "spacing."

Appendix D

Analysis of Individual Test Items

Baseline Comparisons

Multi-step items. Although the overall analysis found no difference between the Soldiers in the dL and traditional classes, performance on the individual multi-step procedures showed some differences (Table D-1). Of the thirteen procedures, group differences were observed on five. Soldiers in the traditional classes scored higher on three of the multi-step procedures: *Create a message folder* (F [1, 212] = 11.33, $p < .01$); *Clear logs and cues* (F [1, 212] = 62.63, $p < .01$); and *Display a satellite image* (F [1, 211] = 4.39, $p < .05$). The dL students performed better on two procedures: *Save an overlay object group* (F (1, 211) = 18.11, $p < .01$) and *Display gridlines on map* (F [1, 211] = 15.01, $p < .01$).

Because there were no systematic differences in performance on these procedures (i.e., sometimes Soldiers in the dL classes performed better and sometimes Soldiers in the traditional classes performed better), it is possible that these differences reflect the different emphasis given by instructors to training on various procedures. Although, the instructors follow the same general program of instruction, individual instructors are free to use whatever training approaches, techniques or examples they see as beneficial (Tucker, et al., 2009).

Operations screen items. Comparisons between the traditional and dL classes also were made for each operations screen question separately (Table D-2). These were single-step items, so the performance measure used was the percent of the sample answering each question correctly. These data were analyzed using the chi square statistic. Of the 14 questions, Soldiers in the traditional classes scored higher on seven questions than the Soldiers in the dL classes (Table D-2): *Enter the MEDEVAC call sign and voice net frequency*, $\chi^2(1) = 17.50$, $p < .01$; *Create a periodic reminder* $\chi^2(1) = 18.40$, $p < .01$; *Assign a message to a quick send button*, $\chi^2(1) = 5.58$, $p < .05$; *Create and save a new position report*, $\chi^2(1) = 19.59$, $p < .01$; *Activate the driver's display for a route*,($\chi^2(1) = 28.36$, $p < .01$; *Display an overlay message that was sent to you*, $\chi^2(1) = 5.49$, $p < .05$; *Transmit your combat platform status/SITREP*, $\chi^2(1) = 14.28$, $p < .01$.

As with the multi-step procedures, differences across training environments may be due to differences in the training emphasis of the instructors. Prior research indicated that the instructors for the traditional classes had more operational experience with FBCB2 (cf. Leibrecht, Goodwin, Wampler, & Dyer, 2007; Tucker et al., 2009). Based on this experience, the instructors of the traditional classes may have provided more operational examples throughout the course that allowed Soldiers to better remember the first steps of these procedures.

Table D-1
Baseline Results for Multi-step Procedures

Multi-step Items – Baseline	Average % Correct *Rank ordered by dL scores*		% Difference (only presented when significant)
	dL (n = 134)	Traditional (n = 80)	
Create, save, and send a SPOT Report	88%	89%	
Set default addressing for a SPOT Report	79%	79%	
Display MGRS gridlines on your map	77%	62%	15% (dL Higher)
Clear logs and queues	77%	94%	17% (Traditional Higher)
Create and send a route	76%	75%	
Create a message folder	70%	82%	12% (Traditional Higher)
Set screen to display all enemy units and only current, friendly units	69%	66%	
Create and save a Combat Services Support overlay	68%	69%	
Center your display on a vehicle	62%	64%	
Display a satellite image on your SA display	58%	68%	10% (Traditional Higher)
Attach an overlay to an OPORD	54%	51%	
Create and Save an overlay object group	50%	38%	12% (dL Higher)

Note. Parentheses indicate group with significantly higher average percent of steps performed correctly. SPOT – includes size, activity, location, terrain; MGRS – military grid reference system; SA – situation awareness; OPORD – operations order.

Table D-2
Baseline Results for the Operations Screen Items

Operations Screen Items - Baseline What is the First Step?	Average % Correct *Rank ordered by dL scores*		% Difference (only presented when significant)
	dL (n = 134)	Traditional (n = 80)	
Create and send an NBC 1 report	98%	99%	
Create and send a Mayday report	96%	99%	
Create an address group	84%	89%	
Assign a message to a quick send button	81%	93%	12%
Create and save a new position report	72%	96%	24%
Display an overlay message that was sent to you	71%	85%	14%
Manually place your icon / platform on the map	69%	59%	
Create a periodic reminder	65%	91%	26%
Use the circular line of sight tool	53%	66%	
Activate the driver's display for a route	49%	85%	36%
Transmit your combat platform status / SITREP	40%	66%	26%
Enter the MEDEVAC call sign and voice net frequency	34%	64%	30%
Check software to troubleshoot a GPS problem	28%	26%	
Edit a location folder	22%	30%	

Note: NBC – nuclear, biological, chemical; SITREP – situation report; MEDEVAC – medical evacuation; GPS – global positioning system

Retention Comparisons

Multi-step items. Performance on each procedure was analyzed with two-factor, time by instructional environment, mixed design ANOVAs which revealed a similar pattern of skill decay for both the dL and traditional classes (see Table D-3). Specifically, there was significant skill decay for 5 out of the 12 procedures for both types of classes: *Create a message folder*, Wilks' $\Lambda = .76$, $F(1, 59) = 18.69$, $p < .01$; *Set default addressing for a SPOT Report*, Wilks' $\Lambda = .90$, $F(1, 58) = 6.14$, $p < .05$; *Clear logs and cues*, Wilks' $\Lambda = .77$, $F(1, 58) = 17.54$, $p < .01$; *Set filters to display all enemy units and only current, friendly units*, Wilks' $\Lambda = .93$, $F(1, 55) = 4.23$, $p < .05$; *Create and save a Combat Services Support overlay indicating the location of a bulldozer*, Wilks' $\Lambda = .78$, $F(1, 52) = 14.78$, $p < .01$.

The responses for the two instructional environments changed differently over time for only one item: *Create and save an overlay object group* as indicated by a significant interactive effect, $F(1, 54) = 9.9$, $p < .01$. For this item, the Soldiers in the dL classes performed the procedure better than the Soldiers in the traditional classes at baseline, $F(1, 54) = 22.7$, $p < .01$, but not at the 8-week measurement. Examination of the means shows that this interaction is primarily due to the poor performance at baseline (35% correct) of the traditional students. Though neither group performed this procedure at particularly high levels at baseline, it would be inappropriate to interpret these findings to mean that the traditional sample retained this procedure better than the dL sample. The better interpretation would be that Soldiers in the traditional classes never learned this procedure to begin with (though it was covered in the course).

It is important to note that for several of the multi-step procedures, performance at baseline was low, thus the lack of significant declines reflects acquisition rather than retention issues. For example, for displaying a satellite image on the map or for attaching an overlay to an operations order, students could only perform 50% to 60% of the steps at baseline and virtually the same percent 8 weeks later.

For the remaining procedures, the number of steps does not appear to be a major factor contributing to skill decay. The number of steps of the retained procedures ranged from 4 – 11. The range of steps of procedures with significant skill decay was 5 – 11.

Operations Screen Items. To better understand the operations screen questions for which Soldiers' performance decreased the most over time, Table D-4 presents the average baseline and retention scores for each question. Wilcoxon signed rank tests were used to assess the significance of changes across the two time points for each instructional environment. Table D-4 shows that for the dL classes, performance on 5 of the 14 questions significantly decreased while for the traditional classes performance on 6 of the 14 questions significantly decreased. It is important to note that four of the questions with significant performance decrements were the same for both types of instructional environments (*Manually place your icon / platform on the map*; *Create a periodic reminder*; *Create an address group*; *Use the circular line of sight tool*).

Table D-3
Retention Results for Multi-step Procedures

Multi-step Items Retention Sample Only	dL *Rank ordered by dL Retention scores*		Traditional		Average % Decrease (only presented when significant)
	Baseline	Retention	Baseline	Retention	
Create, save, and send a SPOT Report	93%	86%	83%	80%	
Display MGRS gridlines on your map	90%	80%	57%	60%	
Set default addressing for a SPOT Report	84%	75%	77%	71%	8%
Create and send a route	81%	74%	72%	67%	
Clear logs and queues	84%	72%	90%	84%	9%
Center your display on a vehicle	74%	69%	57%	53%	
Set screen to display all enemy units and only current, friendly units	82%	65%	61%	58%	10%
Create and save a Combat Services Support overlay	78%	66%	66%	53%	13%
Display a satellite image on your SA display	65%	67%	61%	60%	
Create a message folder	74%	62%	79%	59%	16%
Attach an overlay to an OPORD	59%	60%	48%	56%	
Create and save an overlay object group*	58%	44%	35%	38%	

Note. SPOT – includes size, activity, location, terrain; MGRS – military grid reference system; SA – situation awareness; OPORD – operations order.
* indicates significant interactive effect.

Table D-4

Percent Correct Scores for All Operations Screen Items – Retention Sample Only

Operations Screen Items What is the First Step?	dL *Rank ordered by dL Retention scores*		Traditional		Average % Decrease (only presented when significant)
	Baseline	Retention	Baseline	Retention	
Create and send an NBC 1 report	97%	100%	97%	90%	
Create and send a Mayday report	94%	97%	100%	90%	
Assign a message to a quick send button	78%	77%	90%	90%	
Create and save a new position report	78%	72%	97%	84%	
Create an address group	91%[1]	59%[1]	84%[2]	58%[2]	29%
Display an overlay message that was sent to you	91%[1]	57%[1]	87%	74%	
Use the circular line of sight tool	78%[1]	48%[1]	74%[2]	39%[2]	33%
Manually place your icon / platform on the map	75%[1]	47%[1]	77%[2]	36%[2]	35%
Transmit your combat platform status / SITREP	50%	47%	61%	55%	
Activate the driver's display for a route	41%	38%	84%[2]	39%[2]	
Create a periodic reminder	75%[1]	34%[1]	94%[2]	39%[2]	48%
Enter the MEDEVAC call sign and voice net frequency	34%	22%	61%[2]	13%[2]	
Check software to troubleshoot a GPS problem	28%	22%	29%	13%	
Edit a location folder	25%	13%	26%	23%	

Notes. [1] = Significant performance decrements for Soldiers in the dL classes; [2] = Significant performance decrements for Soldiers in the traditional classes. NBC- nuclear, biological, chemical; SITREP – situation report; MEDEVAC – medical evacuation; GPS – global positioning system.

As mentioned in the prior section, low levels of proficiency at baseline (e.g., 50% or less of the sample peformed the question correctly) followed by similar levels of performance eight weeks later is better interpreted as an acquisition problem rather than skill decay. For example, for the *Enter the MEDEVAC call sign and voice net frequency* and the *Activate the driver's display for a route* questions, the performance of the Soldiers in the dL classes was low and showed a non-significant change over time, while the Soldiers' performance in the traditional classes was high and significantly declined over time.. On the other hand, both dL and traditional students were at relatively high levels of proficiency at baseline for the *Display an overlay message that was sent to you* procedure. The traditional students' performance showed no significant declines over time, suggesting that these Soldiers retained this procedure better than Soldiers in the dL classes whose performance significantly declined (see Table D-4). Wilcoxon Z statistics are presented in Table D-5.

Table D-5
Wilcoxon Signed Rank Scores for the Operations Screen Items.

Operations Screen Items What is the First Step?	*Wilcoxon Z*	
	dL	Traditional
Create and send an NBC 1 report	0.0	-1.0
Create and send a Mayday report	0.0	-1.7
Assign a message to a quick send button	0.0	0.0
Create and save a new position report	-0.8	-1.6
Create an address group	-3.0*	-2.5*
Display an overlay message that was sent to you	-3.1*	-1.6
Use the circular line of sight tool	-2.2*	-3.1*
Manually place your icon / platform on the map	-2.5*	-3.2*
Transmit your combat platform status / SITREP	-0.6	-0.7
Activate the driver's display for a route	-0.3	-3.3*
Create a periodic reminder	-3.2*	-4.1*
Enter the MEDEVAC call sign and voice net frequency	-1.2	-3.9*
Check software to troubleshoot a GPS problem	-0.8	-1.7
Edit a location folder	-1.4	-0.3

* *p* . < 05

Appendix E

Rater agreement and Recall by Soldiers

Another way to summarize the ratings of the non-FBCB2 users is to look at the mean recall rate of items by the number of agreeing raters. As Table E-1 shows, for items in which all six raters agreed on the correct response (15 agreeing pairs), 87% of the Soldiers were able to recall the correct response. In contrast, for items in which none of the raters agreed on the correct response, the recall rate by Soldiers was only 54%. As described above, all Soldiers in these samples correctly responded to these items at baseline.

Mis-cueing by the system is evident when looking at recall rates for items in which raters consistently selected a response that was incorrect. When 5 of the raters agreed (10 agreeing pairs) on a response that was incorrect, the mean recall rate by Soldiers was only 42%. In contrast, when there was no consensus on an incorrect response, the recall rate was 77%.

Table E-1

The Relationship Between Consensus of non-FBCB2 users and Recall by Soldiers

	Recall Rate by Soldiers	
Number of agreeing pairs of naïve raters	naïve raters agreed on correct response(s)	naïve raters agreed on incorrect response(s)
15	87%	--
10	80%	42%
7	82%	--
6	69%	64%
4	88%	63%
3	63%	64%
2	--	59%
1	65%	--
0	54%	77%